

# INTERNATIONAL ENCYCLOPEDIA OF PUBLIC POLICY

## VOLUME 2—ECONOMIC POLICY

EDITOR: PHILLIP ANTHONY O'HARA

GPERU, PERTH  
AUSTRALIA

2011

First published 2011  
by GPERU

*International Encyclopedia of Public Policy:*  
*Volume 2: Economic Policy*

*GPERU is an imprint of the*  
*Global Political Economy Research Unit*

© 2011 Editorial matter and selection, Phillip Anthony O'Hara;  
Individual chapters, the contributors

Typeset in Times New Roman, Algerian, Comic Sans MS  
by GPERU, Perth, Australia.

All rights reserved. No part of this book may be  
commercially reprinted or reproduced or used in any other  
form or by electronic, mechanical or other means, including  
photocopying and recording, or any other information storage,  
without permission by the publisher and editor. Non-commercial use of  
materials by individuals, libraries, universities and governments  
requires proper detailed reference and statement of  
access details of the encyclopedia.

*British Library Cataloguing in Publication Data*  
A Catalogue record for this book is available from the British Library

*Library of Congress Cataloging-in-Publication Data*  
A Catalogue record for this book is available from the Library of Congress

ISBN-13: 978-0-646-97284-8

*For reference only:* ISBN-10: 0-646-97284-7

Libraries Australia ID: 62285981

National Library of Australia Bookmark: <https://trove.nla.gov.au/version/254124757>

# INTERNATIONAL ENCYCLOPEDIA OF PUBLIC POLICY

## VOLUME 2: ECONOMIC POLICY

### EDITORIAL TEAM

#### EDITOR:

*Phillip O'Hara* Global Political Economy Research Unit,  
Curtin University, Perth, Australia

#### ASSOCIATE EDITORS:

*John Burgess* University of Newcastle, Australia  
*Sergio Rossi* University of Fribourg, Switzerland.

#### EDITORIAL BOARD:

*Canir Bakir* Koç University, Istanbul, Turkey  
*Chris Moore* Massey University, New Zealand  
*Reynold Nesiba* Augustana College, South Dakota  
*Jan Reijnders* Utrecht School of Economics, Netherlands  
*Matias Vernengo* University of Utah, Salt Lake City  
*Oren M. Levin-Waldman* Metropolitan College of New York, New York

#### How to reference (example):

How to Reference (eg): Malcolm Sawyer, "Budget Deficits", in P.A. O'Hara (Ed), *International Encyclopedia of Public Policy. Volume 2: Economic Policy*. GPERU: Perth, pp. 41-52. <http://pohara.homestead.com/Encyclopedia/Volume-2.pdf>

Correspondence with Editor: [iepp1@runbox.com](mailto:iepp1@runbox.com)

## Contents of Volume Two

<u>Automatic and Discretionary Stabilizers</u>	<i>John Lodewijks</i>	1
<u>Balance of Payments</u>	<i>Matias Vernengo</i>	16
<u>Bond Market</u>	<i>Caner Bakir</i>	28
<u>Budget Deficits</u>	<i>Malcolm Sawyer</i>	41
<u>Business Cycles</u>	<i>Jan Reijnders</i>	53
<u>Business Cycles: Endogenous Heterodox Perspectives</u>	<i>Howard J. Sherman</i>	77
<u>Central Banking</u>	<i>Chris Moore</i>	86
<u>Chartalism and Metallism</u>	<i>Sergio Rossi and Szymon Klimaszewski</i>	99
<u>Cost Benefit Analysis</u>	<i>Margaret Giles</i>	113
<u>Debt Deflation</u>	<i>Reynold Nesiba</i>	125
<u>Deposit Insurance</u>	<i>Kam Hon Chu</i>	132
<u>Development Governance</u>	<i>Richard Peet</i>	141
<u>Economic and Financial Crises</u>	<i>Christian E. Weller</i>	154
<u>Economic Growth</u>	<i>Amitava Krishna Dutt</i>	165
<u>Employer of Last Resort Schemes</u>	<i>Fadhel Kaboob</i>	179
<u>Employment Agencies and Temporary Work</u>	<i>Julia Connell</i>	194
<u>Endogenous and Exogenous Money and Credit</u>	<i>Sergio Rossi</i>	204
<u>Financial Regulation and Deregulation</u>	<i>S.D. Mavroudeas &amp; D. Papadatos</i>	215
<u>Fiscal Federalism</u>	<i>Duc Hong Vo</i>	230
<u>Fiscal Policy</u>	<i>Leonardo Vera</i>	241
<u>Fiscal Sociology</u>	<i>Michael McLure</i>	256
<u>Free Banking</u>	<i>Kam Hon Chu</i>	265
<u>Free Trade and Protection: Theory and Perspectives</u>	<i>Turan Subasat</i>	278
<u>Future of Work</u>	<i>John Burgess</i>	293
<u>Game Theory</u>	<i>Roger Ashton McCain</i>	301
<u>Global Development and Financial Institutions</u>	<i>Richard Peet</i>	316
<u>Global Policy Coordination</u>	<i>Bin Zhou</i>	326
<u>Government Debt</u>	<i>Matias Vernengo</i>	346
<u>Human Capital and Knowledge</u>	<i>Brian Chi-ang Lin</i>	358
<u>Income Inequality and the Distribution of Power</u>	<i>Oren M. Levin-Waldman</i>	374
<u>Inflation</u>	<i>Sergio Ross</i>	391
<u>Islamic Banking</u>	<i>Feisal Khan</i>	407
<u>Labor Conditions, Hours and Rights</u>	<i>John Burgess</i>	421
<u>Labor Market Liberalization</u>	<i>Mehmet Odekon</i>	434
<u>Labor Market Training Programs</u>	<i>Eduardo Tomé</i>	442
<u>Law and Economics</u>	<i>Natasa Golubovic</i>	455

<b><u>Lender of Last Resort</u></b> <i>Chris Moore</i>	465
<b><u>Liquidity Trap</u></b> <i>Jan Reijnders</i>	482
<b><u>Living Wage</u></b> <i>Oren M. Levin-Waldman</i>	495
<b><u>Long Waves of Growth and Development</u></b> <i>Vincent Barnett</i>	509
<b><u>Moral Hazard and Adverse Selection</u></b> <i>Jong-Shin Wei</i>	521
<b><u>National Accounts</u></b> <i>Yngve Abrahamsen and Jochen Hartwig</i>	527
<b><u>Pareto Optimality, Second Best and Social Welfare</u></b> <i>Michael McLure</i>	544
<b><u>Policy Ineffectiveness Proposition</u></b> <i>Jan Reijnders</i>	555
<b><u>Political Business Cycles</u></b> <i>Jerry Courvisanos</i>	566
<b><u>Productivity</u></b> <i>Gianluca Carnabuci</i>	574
<b><u>Public Goods, Externalities, and Governance</u></b> <i>Jouni Paavola</i>	586
<b><u>Reserve Bank Governor or Chair</u></b> <i>Caner Bakir</i>	598
<b><u>Sickness, Absenteeism and Insurance</u></b> <i>Daniela Andren</i>	608
<b><u>Taxation Compliance Costs</u></b> <i>Jeff Pope</i>	614
<b><u>Taxes-Drive-Money Approach</u></b> <i>Sergio Rossi</i>	627
<b><u>Unemployment Compensation</u></b> <i>Oren M. Levin-Waldman</i>	641
<b><u>Workplace Agreements</u></b> <i>John Burgess</i>	658

## Automatic and Discretionary Stabilizers

*John Lodewijks*

“Automatic stabilizers are a very old idea. Indeed, they are a very old, very Keynesian idea. At the same time, they fit well with the current mistrust of discretionary policy and the focus on policy rules. Should we use them more? Less? Differently? It is indeed a good time to revisit the issues.” *Olivier Blanchard* (2000)

### **Introduction**

Since World War II economic recessions have become less frequent and severe and expansions have lasted longer. Macroeconomic stabilization policy, particularly the use of fiscal and monetary tools, is often credited for playing a key role in these developments (Bell 2000 pp.157-158). Fiscal policy can contribute to stability by smoothing economic fluctuations. It can stabilize output, incomes and demand during an economic downturn by increasing government expenditures or reducing tax revenues. These fiscal stabilizers will cushion cyclical fluctuations.

Fiscal stabilization can result from *discretionary policy-making, where governments actively decide to increase spending or lower taxes*. In contrast, government revenues and expenditures will also change independent of any discretionary action—via the automatic fiscal stabilizers. *Automatic fiscal stabilizers are the reaction of the budget to economic fluctuations in the absence of any government action*. These result from structural features of taxation and social transfers that are built into tax codes and social legislation.

If economic activity weakens, then taxation revenue will automatically decline (as the community's income falls) while government spending would increase (as more people are eligible for unemployment benefits and other welfare support). This automatic fiscal stimulus, associated with the budget moving into deficit (or the surplus falling), cushions the slowdown in economic activity. Automatic stabilizers cause the budgetary balance to move so that it acts counter-cyclically. Normally the larger the public sector, the larger the automatic stabilizers and the more effectively the economy is shielded from economic fluctuations.

In recent times there have been severe doubts raised about the effectiveness of discretionary fiscal fine-tuning as a way of achieving cyclical smoothing. Conversely, the automatic stabilizers have been elevated in status and are judged to be fully consistent with fiscal sustainability. A 2002 report by the European Central Bank illustrates this apparent consensus. The report notes that discretionary policy performed poorly in the euro area from the mid-1970s to the mid-1990s: ‘Discretionary fiscal policies are normally not suitable for short-term demand management ... Past attempts to manage aggregate demand through discretionary fiscal policy-making—or fiscal fine-tuning—have been widespread but often counterproductive’ (ECB 2002:36-38). Budget deficits in downturns were not matched by surpluses in booms leading to a continuous build-up of public debt ratios. It proved politically easier to use expansionary rather than contractionary policies—it was always more popular to cut taxes than to raise them—so there was a bias for continuing deficits. Discretionary reaction by policy-makers proved inflexible, with

long implementation lags, so that when policies finally had an effect, the economic circumstances had often changed. Fiscal policy then could become in effect a procyclical measure, exacerbating instability, rather than a stabilizing element.

By contrast, automatic stabilizers were judged to operate symmetrically over the cycle. It was calculated that in the euro area a one percentage point increase (or decrease) in real GDP growth improves (or worsens) the budget balance by 0.5 percentage point of GDP owing to the operation of the automatic stabilizers. (This average varies across countries depending on their tax structures, welfare programs and size of governments). These automatic stabilizers were directly linked to the structure of the economy and so can respond quickly and transparently. The ECB concluded that automatic stabilizers are all that are normally required to stabilize economies: 'Automatic stabilizers are the appropriate way to stabilize output, as they have foreseeable, timely and symmetrical effects' and 'there is normally no need to engage in additional discretionary fiscal policy-making for stabilization purposes' (ECB 2002:41, 46).

These views led to the fiscal aims embodied in the Maastricht Treaty and the rationale of the Stability and Growth Pact for fiscal discipline. The fiscal directives now involved a budgetary position of balance or surplus in the medium term (over the business cycle). In any particular year the budget balances may improve or worsen in line with cyclical influences (due to the automatic stabilizers) as long as it does not exceed a 3 percent of GDP limit (except in exceptional circumstances like a fall in real GDP of at least 2 percent) under threat of sanctions. Generally, most fluctuations were to be handled by the automatic stabilizers

and only very severe shocks to the economy, beyond the capacity of the automatic stabilizers to dampen, would allow the operation of the discretionary stabilizers. The collapse of these fiscal rules in late 2003, with France and Germany withdrawing their support, indicates that these issues are far from settled in the policy arena.

### **Evolution of Fiscal Stabilizers**

The use of fiscal stabilizers is associated with John Maynard Keynes and the Keynesian policy 'revolution'. In a nutshell, the argument was that a market economy was vulnerable to fluctuations associated with the business cycle and that this economic instability could become particularly severe leading to deep recessions (or even depressions) with painful consequences for unemployment and national income. Hence it was the duty and responsibility for governments to engage in countercyclical macroeconomic management to iron out these fluctuations through activist discretionary fiscal policy, aided by the operation of the automatic stabilizers.

However, the rapidity and extent of the Keynesian conversion varied enormously across countries. If we take the case of the United States, for example, we find that the acceptance of the discretionary stabilizers was long delayed. When reading the macroeconomics literature of the 1940s and 1950s one is struck by the continual emphasis on automatic budget stabilizers and the need for built-in flexibility in the stabilization process. The use of discretionary stabilizers was actively discouraged.

Herbert Stein (1984:79-80) has argued that the reliance on automatic stabilizers,

and only in extreme circumstances deliberate changes in government taxes and spending, was the standard approach to fiscal policy in the Truman and Eisenhower years. The Committee for Economic Development (CED), a private organization of businessmen established in 1942, epitomized the spirit and substance of the postwar fiscal consensus, said Stein. The CED focal 1947 document, entitled 'Taxes and the Budget: A Program for Prosperity in a Free Economy', argued strongly for the automation of fiscal policy supplemented by discretionary action only in serious situations. They advocated confining budget surpluses and deficits to those automatically generated by fluctuations in the level of national income. The structural budget (calculated at levels near full employment) was to be balanced. This was seen as a way to bypass the problems of accurate economic forecasting and the political biases in the implementation of discretionary fiscal policy. Walter Heller in 1957 noted that the 1947 CED statement represented the 'dominant theme in postwar fiscal-policy thinking'.

Economists associated with the University of Chicago were particularly prominent in the assault on the over-optimism of Keynesians with regard to making accurate estimates and forecasts required for 'fine-tuning' combined with their naiveté with respect to the political management of aggregate demand. They argued that a political bias would lead to secular increases in government spending. Henry Simons as early as 1936 had called for a repudiation of discretionary authority. However, Milton Friedman in 1948 with his 'A Monetary and Fiscal Framework for Economic Stability' set the tone for the Chicago approach. Like Simons, Friedman

argued that the government needed to eliminate discretionary authority with respect to monetary and fiscal policy. For stabilization purposes, the economy would have to rely on the automatic budget stabilizers alone. Governments needed the disciplining forces of rules of conduct for policy—like a balanced structural budget—as protection against uncontrollable expenditures and continuous budget deficits.

The mainstream of economic thinking in the U.S. had turned towards automation of fiscal policy. Notions of compensatory or functional finance had been overshadowed by the appeals of automatic flexibility. Walter Heller himself acknowledges that he had been a 'worshipper at the shrine of fiscal automaticity'. Intellectual antecedents for the built-in flexibility view could be found in the writings of Gunnar Myrdal, A.G. Hart, Alan Sweezy, Beardsley Rummler and Alvin Hansen in the late thirties and early forties. After the 1947 CED paper, A.G. Hart, Richard Musgrave, Merton Miller and Milton Friedman further developed the underlying concepts. In 1949 16 economists including Howard Ellis, John K. Galbraith, Paul Samuelson, Sumner Slichter, Jacob Viner and Arthur Smithies signed the 'Princeton Manifesto' supporting policies based on built-in flexibility and automatic adjustments.

Support for policy activism and fine-tuning only really became acceptable in the sixties. In 1957 Walter Heller presented a wide-ranging critique of the CED reliance on automatic stabilizers saying that an increased reliance on discretionary, rather than automatic, stabilizers would improve economic stabilization. Macroeconometric simulations showed that the automatic stabilizers were not strong enough to return the economy to full employment during



recessionary periods. Hence, activist discretionary policy had to bear the brunt of alleviating any significant fluctuations. In 1962 President Kennedy proposed a 'shelf of public works', a standby program of public capital expenditures which could be triggered by increases in unemployment. He also proposed that the benefit period of unemployment insurance be lengthened during times of high unemployment. These measures would strengthen the automatic stabilizers but the key stabilization tool was the use of discretionary policy. Paul Samuelson, reflecting this perspective, argued that optimal policy-making involved continuous alterations in government expenditures and taxation in response to changing circumstances through conscious discretionary policy decisions. There should be no adherence to fixed rules of conduct for policy-making.

We see then that for the United States the desirability of using discretionary stabilizers as the main policy weapon, with the automatic stabilizers playing a supportive role, only became fashionable in the 1960s. This development would not last long; it was barely a decade before it was again fashionable to castigate the discretionary behaviour of governments.

A similar story to the United States experience can be told for Europe. Endres & Fleming (2002) document the evolution of thinking about stabilization policy in key international organizations, many of which were based in Geneva, such as the International Labour Office, Bank of International Settlements and the League of Nations. The discussion in the 1940s illustrated a skeptical perspective on Keynesian countercyclical economic management in these international organizations. There were explicit

reservations about 'pump priming' public expenditure policies and concerns about the crowding out of private expenditures and issue of credibility. Fiscal policy will be ineffective if public opinion did not find government responses credible. Instead of discretionary stabilization policies, governments were instructed to follow simple fiscal policy rules. However, there was general support for strengthening the automatic stabilizers. One idea was to schedule public works according to an index of the intensity of unemployment. A shelf of public works projects would be activated if unemployment reached a certain level (Endres & Fleming 2002:94, 195-196).

The United Nations Committee of Experts Report was published in 1950 and it recommended compensatory schemes, using a range of signals to bring automatically into action, such as differential tax changes. These involved pre-announced rules that automatically came into operation in clearly defined eventualities. There were questions raised about which indicators to use—such as unemployment rates—and the appropriate trigger index. However, what was clear was the preference for rules over discretion. The support for automatic compensatory fiscal policy illustrated the lack of faith in discretionary action. There was credibility attached to a transparent rules-based policy regime (like balanced budgets over the cycle). Fiscal rules were seen as necessary as politicians seemed unable to wisely apply discretionary policy. Even though Keynesian policy advice is often associated with active discretionary management of the domestic economy, by 1950 a predominantly rule-based Keynesianism was articulated by United Nations researchers (Endres & Fleming 2002:224, 229, 235).

## **Discretionary Stabilization Policy Demise**

The traditional Keynesian view of the 1960s and early 1970s saw discretionary fiscal policy as an effective tool for cyclical stabilization. This view collapsed in the second half of the 1970s in an era of declining economic performance and rapid accumulation of public debts. Indeed it can be argued that discretionary countercyclical fiscal policy were last tried in the United States in the 1960s and 1970s and have really been non-operational thereafter. A range of factors led to the abandonment of discretionary stabilizers.

Theoretical and empirical developments in the economics profession played a role. Milton Friedman told us more than 50 years ago to abolish discretionary fiscal activism. His views became far more fashionable. Post-war studies of the United States seemed to show that at best fiscal policy was only weakly stabilizing. New theoretical concepts such as 'Ricardian equivalence' were used to show that that deficit spending was ineffective for stimulating output, as consumers anticipated future tax increases to pay the ensuing public debt, and cut back on consumption accordingly. In turn, contractionary fiscal policy may in fact be expansionary due to effects on financial markets and saving behaviour. It should be noted that the empirical evidence is not supportive of the Ricardian equivalence proposition (Bell 2000:162).

Robert Lucas argued that once expectations of the future were 'rational' the economic actors would anticipate policy effects and take action now that might well nullify the intentions of fiscal policy – the policy ineffectiveness proposition. Indeed Lucas (2003) has gone further and questioned the whole rationale of stabilization policy. The standard view is

that stabilization is an objective of public policy that tries to minimize output volatility by cushioning the economy against temporary shocks to employment, investment, interest rates and the exchange rate. Lucas asks: Is stabilization desirable? His answer is that perfect stabilization would yield virtually no utility gain to households both in absolute terms and relative to the utility gains from a modest increase in economic growth. Many traditional economists are shocked by this perspective and point to the deleterious effects of business cycles on social cohesion and political stability, the negative affect on the minority of the work force unemployed, and the loss of human capital with job losses.

Reinforcing these theoretical debates were disappointing stabilization experiences. There was criticism that the high inflation of 1970s was due to excess demand generated by misguided monetary and fiscal policies attempting to lower unemployment below some 'natural rate'. Stabilization policy, it was argued, was in fact the source of economic instability. In the early 1990s an economic slowdown led to an increase in public deficit in a number of European countries. In France and the U.K. a strong increase in private household saving meant that fiscal policy did not stimulate domestic demand despite these budget deficits. Did consumers perceive increases in future tax obligations to pay for these budget deficits and respond by raising saving or was it just precautionary saving in the context of uncertain employment prospects? Moreover, these ballooning budget deficits raised serious concerns about financial sustainability. From 1980 to 1995 public debt in the European Union on average rose by 35 percent of GDP. The public debt/GDP ratio in several countries showed the largest

ever peacetime increases during the last 20 years of the twentieth century. Elsewhere, it was only in 1998 that the federal government of the United States achieved a budget surplus - for the first time in 30 years—and that surplus did not last long. Had there been a serious breakdown of fiscal discipline and how could it be reversed?

Economists increasingly turned to a new sub-discipline of the profession to explain these developments. This field is variously described as public choice theory or constitutional economics or the economics of politics or more broadly as the new political economy (see Cukierman et al (1992), Breton (1998), Drazen (2000), Buti et al (2002), Burger (2003), Pitt et al (2004)).

The new political economy literature views economic policy as the outcome of a process in which distributional considerations and political competition play a key role. It is the study of political incentives and political institutions and how conflicting interests are aggregated into public policy. Two alternative approaches to political decision-making are often contrasted. The first sees policymakers as social planners who devise policy so as to maximize social welfare. This view is regarded as naïve at best. A more cynical approach is to view politicians as self-interested and susceptible to all sorts of political and self-interest pressures. A pioneering application of this idea was originally proposed by Michal Kalecki and called the political business cycle. This hypothesis is that politicians stimulate aggregate demand before elections to create fast growth and reduce unemployment. After their election win the inflationary consequences are eliminated via contractionary policy. There is some

empirical support for this position - preelectoral manipulation of economic policy occurs frequently and fiscal deficits are higher in election years by more than 0.6 percent of GDP (Cukierman et al 1992:228, 252).

Not only are there political cycles in policy instruments, such as soft fiscal policy prior to elections, but the nature of political institutions influence fiscal performance. For example, multiparty coalition governments, especially those with a short expected tenure, are poor at reducing budget deficits. Helpman et al (1988) also document issues of time inconsistency with respect to fiscal policy. Suppose a government currently in power knows that it will eventually be replaced by a new government with different objectives, for example, a government that favours a large public sector. How should it act to influence how a future government governs? It might act by deciding to collect less taxes and leave a sizeable public debt (higher than it would have left if it had remained in power) for its successor. This then constrains the actions of future governments and creates a deficit bias.

Shultz & Dam (1977) argue on the basis of their American policy experiences that it is extremely difficult to make spending cuts as each program has a political clientele that will delay and subvert any cuts. There will be collusion among interest groups—they call them rent-seekers—to capture government departments so that they protect the spending in particular sectors. There is also the political infeasibility of running large surpluses in boom times. These surpluses unleash intense political pressure to increase public spending. Policy is then shaped by electoral and political systems and cannot be flexibly directed to economic management. They conclude that what are

required are institutional restrictions on government such as rules based fiscal policy. Binding fiscal rules are needed to counter political bias and spend-thrift politicians. These rules limit and discipline vested interests. This perspective reinforces the views of Lucas who argues that we need to focus on alternative policy rules. In his view the analysis of policy which utilizes economics in a scientific way necessarily involves choices among stable, predictable policy rules, which are infrequently changed. This would minimize the role of discretionary economic management. Yet will this deprive the policymaker of an important stabilization tool?

### **Fiscal Discipline**

In policy circles we have seen a major retreat from the use of discretionary policy for short-run stabilization purposes. Monetary policy has effectively been taken away from governments and left under the control of independent central banks. These central banks have pursued inflation targeting to rid the economy of an inflationary bias. Similarly, with respect to fiscal policy we have balanced budget propositions and stability pacts aimed at restricting the discretion that policymakers can have. This is intended to isolate fiscal decisions from special interests to avoid a deficit bias. Unless constrained, political and electoral systems are alleged to generate incentives to overuse discretionary fiscal policies (Buti et al 2002).

Both in the United States and Europe the role of fiscal policy has been actively debated. In the United States vigorous discussion about a Balanced Budget Amendment took place. Any such Amendment would make redundant the use of fiscal policy to stabilize business cycle

fluctuations. Nominal caps were imposed on discretionary public spending but these have been considerably eased in recent years. Moreover, the 2003 Bush administration tax cuts are projected over the next 75 years to lead to a loss of taxation revenue three times that of the projected social security shortfalls.

European Union countries have implemented clear fiscal rules. The Stability Pact of the Maastricht Treaty involves budget deficits constrained at 3 percent of GDP and gross public debt constrained at 60 percent of GDP. Budgets are to be balanced or in surplus over the medium term. The Pact rules try to restrict overly active fiscal management and excessive deficits that cannot be justified on the basis of government investment expenditure or economic distress. As a result, OECD countries have reduced their aggregate budget deficits from 5.5% of GDP in 1993 to less than 3% in 1997 and budgets were on average balanced in 2000. However, since then the situation has deteriorated. In 2002 the average government deficit was 3 percent of GDP and the gross public debt/GDP ratio reached 75 percent.

Perry (2003) demonstrates that the Mercosur and Andean countries are also following the EU example and adopting fiscal and debt ceilings. But Latin American and Caribbean countries are much more volatile than their European counterparts (indeed, twice as volatile as OECD countries) and generally apply procyclical fiscal policies that exacerbate volatility. There is evidence of a significant negative effect of macroeconomic volatility on long term growth in the region, but also of a perverse relation between volatility and poverty, education, income distribution and financial deepening. Fiscal policies tend to

be procyclical, so that social expenditures tend to fall during downturns and grow during upturns. Moreover, expansionary fiscal policies during booms have led to crises during downturns.

What explains the procyclical behaviour in this region of the world? These countries have fragile fiscal positions even in boom times for they have a record of excessive expenditures even when the economy is expanding. Political pressures and incentives are aligned with spending out any surplus in a boom. This generates a clear deficit bias and strong default risk. There are serious credibility problem for these countries when they attempt expansionary fiscal policies. Hence it is argued that we need to tie the hands of the discretionary authorities via explicit fiscal rules. These constraints keep a surplus in good times out of reach of the political interests and enhance fiscal credibility.

Recently proposed and adopted Fiscal Responsibility Laws operate in Brazil, Argentina, Peru, Colombia and Ecuador. Chile has an explicit commitment to keep a 1 percent structural surplus each year yet allows full operation of the automatic stabilizers during the economic cycle.

Sub-Saharan Africa (SSA) has an even more desperate situation for government borrowing (World Bank 1989), since long-term debt rose 19-fold in the 1970s and 1980s. SSA is the most heavily indebted region of the world. A significant effort will be required to generate fiscal discipline through promoting public savings via revenues and controlling expenditures. In contrast, macro-management in ASEAN and East Asian countries has been much better than the developing country average. (But see Joshi & Little's (1994) study of India's unsustainable fiscal policies). These

countries have generally run modest budget deficits and yet have been able to supply badly needed public goods such as infrastructure and education. Fiscal policy has also been remarkably flexible in the face of exogenous shocks. The fragmented politics of Africa, and the power of organized vested interests in Latin America, are in marked contrast to the Asian record of prompt and effective fiscal adjustment (World Bank 1993:ch 3; Woo et al 1994).

### **Fiscal Illusion?**

The search for an effective menu of rules for the maintenance of fiscal sustainability and economic stability seems to be at the forefront of current discussions (see Poterba and Hagen (1999), Burger (2003)). Yet attempts to restore fiscal sustainability have not always worked well in practice. William Easterly (1999) provides convincing evidence showing that attempts by governments to reduce the fiscal deficit or public debt may be illusory or counter-productive in practice so that: "Compliance with fiscal rules has led to ... cuts in public investment ... accumulation of payment arrears, proliferation of creative accounting practices and recourse to one-off measures (such as financing from privatization receipts)" (Easterly 1999:62).

Responses to the pressures for fiscal austerity include budget gimmickry such as the sale of public assets. There are short-sighted cuts in public health measures in the name of fiscal austerity which can result in a future "public health bill many times what it failed to spread on prevention, not to mention the loss of life". Another popular, and short-sighted approach is to cut public investment: "in periods of restrictive fiscal policies ... capital expenditures are the first to be reduced (often drastically)" (Easterly

1999:59). During fiscal adjustments governments cut capital expenditures by four times that of current expenditures - this is well documented in both developing and OECD countries. Yet World Bank estimated rates of return to infrastructure investment vary from 19 percent (telecommunications) to 29 percent (highways). So these reductions are a false economy. Another area cut is to reduce operations and maintenance spending and this has resulted in massive future road reconstruction, reduced generating capacity and breakdowns in power plants, water wastage, deteriorating bridges and ports and faulty medical equipment in hospitals. Allowing these assets to deteriorate hurts future government revenue and simply defers greater infrastructure spending to the future.

It also tends to redirect spending to projects that disburse slowly, whatever the long-run costs of the project, or delays addressing problems such as financial crises in banks adding to the cost of the eventual payout by government. Another trick is to require pension funds to lend to the government at negative real interest rates. In Peru the real return on the pension fund was - 37.4 percent. This cuts the fiscal deficit today but also cuts pension fund assets available to meet future claims.

A fiscally irresponsible government can frustrate any attempt to control its finances through constraints on the conventional budget deficit. Visible fiscal adjustment is then an illusion, says Easterly. These findings also apply to Europe. The Maastricht controls encourages financial engineering to avoid underlying real fiscal adjustment: 'European governments have relied on one-off measures—central bank sales of gold, refundable euro taxes, appropriation for the general budget of

public enterprise reserves and sales of strategic petroleum reserves—to meet the Maastricht fiscal criteria' (Easterly 1999:71). Governments have reduced public investment, privatized aggressively and thereby sold valuable state assets, and accumulated implicit but unrecorded liabilities.

Clearly on the basis of this evidence, and the fact that the Fiscal framework of the Growth and Stability Pact has been suspended, with France and Germany rejecting the fiscal rules of the European Monetary Union, we are still some distance from formulating successful fiscal rules. One lesson is that we need to avoid rigid quantitative ceilings. Secondly, targets should be based on structural deficits, excluding cyclical fluctuations, and flexible enough to accommodate changes in societal preferences.

### **Appeal of Automatic Stabilizers**

Automatic stabilizers have survived the critique of fiscal policy much better than the activist policy proposals. Even those that believe that discretionary fiscal policy has been largely discredited as an instrument to fine-tune the economy regard automatic stabilizers as the 'norm' for budgetary behaviour. We have already seen how those from the right of the political spectrum like Milton Friedman have supported these automatic stabilizers. But there is substantial support from those critical of market capitalism. Hyman Minsky (1986) is a case in point. Minsky argued that a market economy was prone to instability and *insisted that the existence of automatic fiscal stabilizers meant the difference between failed and successful capitalism*. The economy needed the stabilizing effects of a big government that can run a sizeable

public deficit or surplus as required. Minsky suggested that outlays and revenues must be of the same order of magnitude as private investment—for the United States this was around 17 percent of GDP—to be an effective stabilizing force. He regarded the unemployment insurance system as the most important stabilizer. A 1999 study confirmed Minsky's view and reported that the loss in GDP would have been 15 percent on average deeper without unemployment insurance (Whalen & Wenger 2002). In contrast, Minsky thought that discretionary action came with long lags that could not be relied upon and so capitalism needed a built-in structure of automatic stabilizers.

Minsky was particularly troubled by the erosion of these automatic stabilizers in the United States over time. With progressive taxation, an expanding economy pushes more and more individuals and families into higher brackets. This raises government revenues automatically and so the growth of aggregate demand is counteracted. A contracting economy does the reverse. But the degree of progressivity has declined with falls in the top tax rate, fewer brackets and brackets indexed to inflation. This erases the stabilizing element of the tax system. Welfare programs based on financial need acted as an automatic stabilizer as payments increased during downturns. But these have been scaled back. Also minimum wages set a floor to wage payments during a downturn but the minimum wage has declined in real terms. There are also now shrinking benefits and limited eligibility for unemployment insurance. The unemployment trigger that extended benefits across the nation during slowdowns has been eliminated. Minsky believed it was vital to shore up the fiscal stabilizers, to reverse the changes that were eroding the automatic stabilizers of their

effectiveness, and to have government as employer of last resort. A guaranteed public sector job would not only be a stabilizing influence but a humane policy response to the vicissitudes of capitalism (see Bell 2000).

Empirical research generally supports Minsky's position on the effectiveness of these automatic stabilizers. One study of 20 OECD countries from 1960 to 1997 shows a clear negative correlation between government size and GDP volatility. A negative correlation between the size of the government budget and the amplitude of business cycles indicates that larger governments stabilize output. In particular, changes in taxes and transfers over the business cycle smooth disposable income. Over the 1990s the automatic fiscal stabilizers worked to dampen cyclical fluctuations by roughly a quarter on average. Another study concluded that fiscal federal stabilizers offset 20-30 percent of any initial reduction in income. Eichengreen (1996) calculates that automatic stabilizers offset about 36 percent of a fall in output generated by a negative shock. In Finland and Denmark without the automatic stabilizers output volatility in the 1990s would have been twice as high (Cohen & Follette 2000). In Canada the cost of suppressing the automatic stabilizers is an increase in GDP volatility by 18 percent and in Germany by 7 percent. Dungan & Murphy (1995) have evaluated the effectiveness of Canada's unemployment insurance system as an automatic stabilizer. They find that it reduces the impacts of shocks to GDP and employment by 15 to 20 percent. Fatas and Mihov (2001) note that a United States Treasury Department study concluded that in the absence of automatic stabilizers would have added 1.5 million people to the ranks

of the unemployed during the last recession. The Federal tax system in that country offsets as much as 8 percent of initial shocks to GDP.

Despite these large benefits there are some alleged costs associated with the automatic stabilizers. A high tax burden and generous social payments reduce incentives to work, invest and innovate. Yet stressing the distortions caused by government policies neglects the redistribution, insurance and revenue generation accomplished by these policies. Automatic stabilizers tend to increase net transfers to the vulnerable households (unemployed) whose propensity to consume is likely to be high. A further complaint is that sizeable automatic stabilizers can delay adjustments in the economy. Automatic stabilizers that stabilize supply shocks may prevent needed adjustments in output. There appears to be a need to develop automatic stabilizers that better cater for supply shocks - such as a proportionate tax on the price of oil. New stabilizers for demand shocks might also be explored—like a tax on layoffs or investment tax credits.

There are also theoretical arguments that place more weight on private rather than public stabilizers. If there are no credit constraints, households can optimally stabilize themselves to maintain constant consumption throughout the cycle without needing government assistance (via borrowing and lending). In such an unrealistic world we might not even distinguish between automatic and discretionary stabilizers. Both lead to future expected tax changes so that the permanent income of households remain unchanged. There would then be no productive role for government fiscal policy at all. But in reality the automatic stabilizers help those

households that are most liquidity constrained and in long run do not increase public debt.

A far more realistic constraint on the operation of automatic stabilizers are the implementation of rigid fiscal rules. The debate on fiscal policy in Europe centers on how to facilitate the workings of automatic stabilizers while achieving fiscal consolidation. Does budget balance refer to the actual budget or the cyclically-adjusted structural budget balance and is the 3 percent of GDP budget limit too severe? If the actual budget is to be balanced then the automatic stabilizers are negated. Note also that fiscally responsible Scandinavian countries have run deficits as high as 8 percent of GDP. Noord (2000) finds that automatic fiscal stabilizers have generally reduced OECD cyclical volatility in the 1990s but in some countries the need to undertake fiscal consolidation has forced governments to take discretionary actions that have reduced if not offset the beneficial effects of the automatic stabilizers. He finds in the 1990s that discretionary fiscal policy acted as a powerful complement to automatic fiscal stabilization in OECD but in the European Union the tight stance of discretionary fiscal policy contributed to the sluggish economic performance.

The Stability Pact with its adherence to the Maastricht restrictions has been strongly criticized for not leaving enough scope for the operation of the automatic stabilizers. Governments trying to bring down public deficits and public debt to comply with the convergence requirements have generated contractionary effects on the economy and subsequently output and income growth in Europe has fallen. Similarly, a balanced budget amendment in the United States would have the same effect. Indeed in that



country, 49 States operate under statutory or constitutional restrictions on debts and deficits of some sort. Those States with the strictest restrictions stabilize the least. Canadian provinces have considerable fiscal freedom to operate their own counter-cyclical policies unlike these American States which are encumbered by balanced budget rules that create fiscal inflexibility.

The European Union experience leads to a broader policy concern. With the creation of a single currency and the disappearance of national monetary policies, the debate has focused on the need for fiscal federation. In Europe, monetary unification increases the need for national automatic fiscal stabilizers. National monetary policies are no longer available to cushion shocks. The exchange rate has already been lost as an independent policy instrument. There is a European Central Bank for monetary policy. Should we establish a central EU authority for fiscal policy? Would fiscal stabilization at an EU-wide level be more effective than devolved to nations? The debate in Europe is about the lack of fiscal stabilization at the EU level, in contrast to the existence of important fiscal stabilizers at the federal level in other monetary unions.

Supporters argue that Europe needs a fiscal federation or at least closer fiscal coordination. With the loss of monetary policy independence, and with fiscal policy constrained by rules, this leaves economies very vulnerable to instability. As prices and wages are not highly flexible, and there is only limited labour mobility, there is the risk of deep recessions and widespread unemployment. A Sustainability Council for the euro area could be established with national governments submitting their fiscal plans to this Council. The Council might develop some sort of an insurance

mechanism to channel income from countries enjoying a positive shock to countries suffering a negative shock, based on growth differentials relative to the EMU average growth rate or by using unemployment based mechanisms. Opponents to this idea claim that such a system would be highly regressive as the faster growing countries are also poorer and that it is not wise to reward a country's poor unemployment record. Eichengreen (1996) argues that a fiscal federation is politically and administratively too hard to achieve and we should just let the automatic stabilizers operate.

### **Future of Fiscal Stabilizers**

Current macroeconomic policy discussion is all about evaluating policy rules. For example, one such simple fiscal rule might be that the actual budget should equal the structural budget surplus plus an adjustment (positive or negative) for any output gap. This allows the automatic stabilizers free rein as they are judged to be more flexible and work more quickly than the discretionary ones. The automatic stabilizers represent a predictable and systematic response, setting out rule-like mechanisms for changes in spending and taxes. Over the last four decades in the United States the role of the automatic stabilizers has been much larger than the role of discretionary fiscal policy. This may change in the future if interest rates remain low. Rates at almost zero imply that monetary policy may no longer be a viable corrective in times of recession. Then there is no alternative but expansionary and discretionary fiscal policy. Also, while increased globalization may constrain domestic policy-making, this may apply more to monetary policy constrained

by increasing capital mobility, than it does to fiscal policy (Bell 2000:169).

One concern is that the size of the automatic stabilizers has declined since the 1980s. Countries have been curbing the effectiveness of the automatic stabilizers as they reduce transfer mechanisms, cut taxes and the number of tax brackets. They are still more sizeable in Europe than in the United States. Yet even in Europe the need to undertake fiscal consolidation has reduced, or even offset, the effect of the automatic fiscal stabilizers. Rigid quantitative budget rules precludes the full extent of the automatic stabilizers.

There is room to strengthen old automatic stabilizers and to develop new ones. Perhaps consumption taxes are efficient stabilizers. For expenditure programs we need temporary stimulus that disappears once the economy has returned to normal such as Federal grants to States based on the amount of excess unemployment. Devising mechanisms yielding temporary stimulus but which would self-terminate as recovery ensued—based on unemployment trigger mechanisms—are a challenge to policy-makers.

Yet there a bigger challenge on the horizon that might dominate discussions of fiscal consolidation. Pension reforms will be the critical fiscal issue of the next decade. In the OECD there has been major increases in transfer spending with the ageing population. Tax revenues cannot keep up with the growth of mandatory entitlement payments. The root of the financial sustainability problem is the increase in the old age dependency ratio. In the EU pension expenditures are projected to increase by 4 percentage points of GDP over 2005-2050 and total ageing-related expenditure by 6 percentage points. Financial problems in the

United States relate to the funding of social security. For this alone a 15 percent increase in payroll tax revenue is required. The share of GDP needed to provide all the benefits scheduled in current law would increase from 4.4 percent of GDP now to 7.0 percent in 2077. There are even larger and more complex problems associated with Medicare and Medicaid. This has severe implications for budget deficits and public debt ratios (Alesina et al 2000; Fatas & Mihov 2003). A World Bank Research Report (1994) titled “Averting the Old Age Crisis” deals with these issues in not only OECD countries but also in Asia, Latin American and Africa.

Finally, two features of the discussion on fiscal stabilizers stand out. One is the difficulty of forecasting and predicting economic instability, taking discretionary action in the public interest and achieving the intended outcome. Ormerod (1998:Ch. 6&7) provides an interesting discussion of this in an evolving, complex system. He suggests that the economic system is extremely difficult to predict and control and governments need but do not have a reasonably exact understanding of what the effects of various policy measures have on the economy. The result is the ‘Law of Unintended Consequences’ where the impact of policy turns out to be the opposite of what is intended, or even if it succeeds in its aims, there are unforeseen adverse consequences elsewhere.

The other remarkable feature of the discussion is the widespread support among economists for the use of automatic stabilizers. Economists as ideologically disparate as Milton Friedman, John K. Galbraith, Hyman Minsky, Gunnar Myrdal and Paul Samuelson have expressed strong adherence to the value and power of these instruments. Given this widespread support

for the role of automatic stabilizers, it is clearly a policy instrument that must be strengthened and can not be allowed to diminish in effectiveness.

### Internet Sites

Centre for Economic Policy Research

[www.cepr.org/pubs](http://www.cepr.org/pubs)

Economic Policy Institute [www.epinet.org](http://www.epinet.org)

Global Policy Network [www.gpn.org](http://www.gpn.org)

Journal of Economic Policy [www.economic-policy.org](http://www.economic-policy.org)

### Selected References

Alesina, Alberto et al. (2000) "Symposia on Fiscal Policy", *Journal of Economic Perspectives*, Volume 14, Number 3, Summer, pp.3-74.

Bell, Stephen. (2000) *The Unemployment Crisis in Australia*. Cambridge: Cambridge University Press.

Breton, Albert. (1998) *Competitive Governments: An Economic Theory of Politics and Public Finance*. Cambridge: Cambridge University Press.

Burger, Philippe. (2003) *Sustainable Fiscal Policy and Economic Stability*. Cheltenham, UK: Edward Elgar.

Buti, Marco, Jurgen von Hagen and Carlos Martinez Mongay. (2002) *The Behaviour of Fiscal Authorities: Stabilization, Growth and Institutions*. New York: Palgrave.

Cohen, Darrel and Glenn Follette. (2000) "The Automatic Fiscal Stabilizers: Quietly Doing Their Thing", *Economic Policy Review*, Federal Reserve Bank of New York, April, with commentary by Olivier Blanchard, pp.35-74.

Cukierman, Alex, Zvi Hercowitz and Leonardo Leiderman. (1992) (Editors) *Political Economy, Growth and Business Cycles*. Cambridge, Mass: MIT Press.

Dungan, Peter and Steven Murphy. (1995) *The Unemployment Insurance System as an Automatic Stabilizer in Canada*. Toronto: Government of Canada, May.

Drazen, Allan. (2000) *Political Economy in Macroeconomics*. Princeton: Princeton University Press.

Easterly, William. (1999) "When is Fiscal Adjustment an Illusion?" *Economic Policy*, Volume 28, April, pp.57-86.

Eichengreen, Barry. (1996) "Saving Europe's Automatic Stabilizers", *National Institute Economic Review*, Volume 159, January, pp. 92-98.

Endres, Anthony M. and Grant A. Fleming. (2002) *International Organizations and the Analysis of Economic Policy, 1919-1950*. Cambridge: Cambridge University Press.

European Central Bank. (2002) "The Operation of Automatic Fiscal Stabilisers in the Euro Area", *Monthly Bulletin*, April pp.33-46.

Fatas, Antonio and Ilian Mihov. (2001) "Government Size and Automatic Stabilizers: International and Intranational Evidence" *Journal of International Economics*, Volume 55, Number 1, October, pp.3-28.

Fatas, Antonio and Ilian Mihov. (2003) "The Case for Restricting Fiscal Policy Discretion" *Quarterly Journal of Economics*, Volume 118, Number 4, November, pp.1419-1447.

Gali, Jordi and Roberto Perotti. (2003) "Fiscal Policy and Monetary Integration in Europe", *Economic Policy*, Volume 18 October pp.535-572.

Hairault, Jean-Olivier, Pierre-Yves Henin and Franck Portier. (1997) *Business Cycles and Macroeconomic Stability: Should We Rebuild Built-in Stabilizers?* Boston: Kluwer.

- Helpman, Elhanan, Assaf Razin and Efraim Sadka. (1988) (Editor) *Economic Effects of the Government Budget*. Cambridge, Mass: MIT Press.
- Joshi, Vijay and Ian M. D. Little. (1994) *India: Macroeconomics and Political Economy*. Bombay: Oxford University Press.
- Lucas, Robert. (2003) "Macroeconomic Priorities" *American Economic Review*, Volume 93, Number 1, pp.1-14.
- Minsky, Hyman P. (1986) *Stabilizing an Unstable Economy*. New Haven: Yale University Press.
- Ormerod, Paul. (1998) *Butterfly Economics*, Faber and Faber, London.
- Pechman, Joseph A. and N.J. Simler. (1982) (Editors) *Economics in the Public Service*. New York: W.W. Norton.
- Persson, Torsten and Guido Tabellini. (2000) *Political Economics: Explaining Economic Policy*. Cambridge, Mass: MIT Press.
- Perry, Guillermo. (2003) "Can Fiscal Rules help Reduce Macroeconomic Volatility in the Latin America and Caribbean Region", *World Bank Policy Research Working Paper*, 3080, June pp.1-24.
- Pitt, Joseph C., Djavad Salehi-Isfahani and Douglas W. Eckel. (2004) (Editors) "The Production and Diffusion of Public Choice Political Economy", *The American Journal of Economics and Sociology*, Special Invited Issue, Volume 63, Number 1, January, pp.1-271.
- Poterba, James M. and Jurgen von Hagen. (1999) (Editors) *Fiscal Institutions and Fiscal Performance*. Chicago: University of Chicago Press.
- Shultz, George P. and Kenneth W. Dam. (1977) *Economic Policy Beyond the Headlines*. New York: W.W. Norton.
- Stein, Herbert. (1969) *The Fiscal Revolution in America*. Chicago: University of Chicago.
- Turgeon, Lynn. (1997) *Bastard Keynesianism: The Evolution of Economic Thinking and Policy-Making Since World War II*. Westport, Conn: Greenwood Press.
- Van den Noord, Paul. (2000) "The Size and Role of Automatic Fiscal Stabilizers in the 1990s and Beyond" *OECD Working Paper* Number 230. Reprinted in Buti et al (2002), pp.130-148.
- Whalen, Charles and Jeffrey Wenger. (2002) "Destabilizing an Unstable Economy: The Market Failures Issue", *Challenge* Nov.-Dec.
- Woo, Wing Thye, Bruce Glassburner and Anwar Nasution. (1994) *Macroeconomic Policies, Crises, and Long-Term Growth in Indonesia*. Washington DC: World Bank.
- World Bank. (1993) *The East Asian Miracle*. Oxford: Oxford University Press.
- World Bank. (1994) *Averting the Old Age Crisis*. Oxford: Oxford University Press.
- World Bank. (1989) *Sub-Saharan Africa: From Crisis to Sustainable Growth*. Washington DC: IBRD.

John Lodewijks  
School of Economics and Finance  
University of Western Sydney, Australia.  
j.lodewijks@uws.edu.au

## Balance of Payments

*Matias Vernengo*

### **Introduction**

The Balance of Payments summarizes all the transactions between a country and the rest of the world. The BP is usually divided into two main accounts, namely: the Current Account (CA) and the Capital Account (KA).

The current account includes the exports and imports of goods and services, the former appearing as credit items and the latter as debit components. Exports of commodities give rise to a claim on the rest of the world that foreigners must discharge by making payments to the domestic producers, and *vice versa* in the case of imports of commodities. Exports and imports of services—such as travels, interest and dividends of investments, and unilateral transfers—imply analogous transactions. It is important to note that interest payments on outstanding debt are part of the current account, and in several cases this is the most important component of the balance of payments. The transactions in the CA generate income flows, and the CA is in equilibrium, surplus or deficit, if payments equal, fall behind or exceed receipts, respectively.

The capital account includes foreign direct investment (FDI) and portfolio investments, in which the latter constitute the so-called hot capital flows, that is, the purchase of financial assets rather than equipment, machines or installations. When a domestic firm, or household, purchases foreign assets—e.g. a productive plant, real state, or a financial instrument—an outflow of capital is generated. Capital outflows are accounted as debits, since the domestic

buyer has to pay to the foreign seller, in the same way that an importer of goods and services would do. By symmetry capital inflows appear as a credit item. The transactions in the KA generate asset flows, and the KA is in equilibrium, surplus or deficit, if payments equal, fall behind or exceed receipts, respectively.

The overall BP is given by the net result of the CA and KA. So that, if a CA surplus is matched by a deficit in the KA, then the BP is in equilibrium. In a fixed exchange rate system—when the monetary authority stands ready to buy and sell the major currencies on a continuous basis, at specified bid and ask prices—an overall BP surplus or deficit may occur. When there is a balance of payments surplus the official exchange reserve holdings of the central bank will increase, and they will decrease in the case of a BP deficit. In formal terms

$$(1) \quad BP = CA + KA = \Delta R$$

where  $\Delta R$  stands for the variation in official reserve holdings. For example, if a CA surplus exceeds a KA deficit, there will be an excess demand for the domestic currency. To avoid the appreciation of the domestic currency the central bank will sell domestic currency, and accumulate foreign reserves.

Under a flexible exchange rate regime—in which the exchange rate is free to float without intervention from the central bank—the overall BP must be in equilibrium, since deficits and surpluses will be eliminated by exchange rate changes, rather than changes in reserve holdings. A surplus in the CA implies that the domestic currency will appreciate, leading to a rise in the price of imports and a fall in the price of exports, that will stimulate exports and

discourage imports, eliminating the CA surplus.

The remainder of this entry will discuss briefly the main theories, and their limitations, explaining the process of balance of payments adjustment, and the literature on the causes of balance of payments crises. It will also analyse the contention that the balance of payments is the main constraint to economic growth. The last section discusses the policy lessons associated with the recent balance of payments experience.

### **Balance of Payments Adjustment**

The theory of balance of payments adjustment has gone in full circle, from the automatic adjustment views of David Hume's specie-flow mechanism, to the Keynesian interventionism of the neoclassical synthesis, to the revival of hands off views within the intertemporal approach.

David Hume (1752) developed the price-specie-flow mechanism not only as an interpretation of the BP adjustment process, but also as an argument against the mercantilist defence of government intervention. According to the price-specie-flow mechanism the BP is self-adjusting. If a country runs a trade deficit, then there will be an outflow of capital, which will lead to deflation in the deficit country, and to inflation in the surplus country. As a result of the fall of prices in the deficit country, its exports will become more competitive, thus restoring the trade balance equilibrium. In other words, capital (gold) flows eliminate any trade imbalance.

The balance of payments adjustment is a purely monetary phenomenon, and all the adjustment is done by changes in relative prices (one must not that Hume himself

admitted short run changes in the level of activity). This was the standard model for balance of payments adjustment by the time of the final collapse of the Gold Standard in the 1930s—and still is in a sense (Eichengreen 1996).

Hume's specie-flow was thought for a world with fixed exchange rates. In the 1930s that assumption became considerably less relevant, and new ideas had to be developed. The elasticities approach was for a while the dominant model, and emphasized the role of substitution effects in bringing the balance of payments to equilibrium.

Most authors at that point remained prisoners of Say's Law, and as a result the level of activity was excluded from any role in adjusting macroeconomic disequilibria. It was only with the formal development of the principle of effective demand by John M. Keynes that the possibility of having the level of income as the adjusting variable entered the scene. As correctly pointed out by Taylor (1990:73), "this [Keynesian] revolution fundamentally attacked Say's Law, and hence the specie flow mechanism."

In Keynes's work the level of income works as the adjusting variable between savings and investment. In an open economy environment the level of income operates as the adjusting variable for a trade deficit (Harrod 1933). That is, if a country runs a persistent trade deficit, and capital inflows are lacking, then a reduction in the level of income would lead to a contraction of imports, and the adjustment of the balance of payments. This came to be known as the absorption approach to the balance of payments. The absorption approach also meant that there was a great degree of elasticity pessimism, that is, the idea that

depreciation would have a minor effect in adjusting the balance of payments.

Structuralist authors pointed out later that even exchange rate movements affect the balance of payments not through its impact on price competitiveness, but through its effect on income distribution and the level of activity. Krugman and Taylor (1978), building on the work by Albert Hirschman and Carlos Diaz-Alejandro, show that depreciation leads to a contractionary adjustment if the economy has a trade deficit or if it redistributes income to higher income groups. In the first case, if the volume of imports is high and the value increases after devaluation, contraction of output may be the only way to reduce the trade deficit. In the second case, if the redistributive effect of depreciation increases the income to low spending groups (higher income groups), then a contraction of output also follows.

The Mundell-Fleming model—a great synthesis of the absorption, elasticities and monetary approaches to the balance of payments developed during the 1950s and 1960s that started with James Meade's (1952) classic *The Balance of Payments*—remained for a long while the dominant view on balance of payment adjustment. In this view, then the adjustment is partially done by changes in the relative prices and partially done by variations in the level of activity. However, to the extent that economists relied more on the concept of a natural rate of unemployment—associated to some optimal level of output—it became evident that in the long run, variations in the level of output cannot be central for balance of payments adjustment.

More importantly, in monetarist criticism of Keynesian models of balance of payments adjustment noted that the latter analysis did not take into consideration the accumulation

of stocks. In other words, Keynesians analyzed the flows of goods, services and capital, but not the accumulated stocks of debt in the form of assets that resulted from balance of payments disequilibria. The monetary approach to the balance of payments and the intertemporal approach that finally came to dominate were designed to solve that problem. In both cases, a central part of the analysis consists on the fact that a country's ability to spend more than it earns is limited by a budget constraint. In the monetary approach the emphasis is on the control of the domestic money supply stock, while the intertemporal approach emphasizes the possibility to smooth out spending patterns over long periods of time. The main conclusion from the intertemporal approach to the balance of payments is that if a country runs a current account deficit in the early periods – for example because it has fiscal deficit and the output level is above the natural level – then it must run a surplus in the future in order to pay the debt that is accumulated in the initial periods (Sachs 1981; Obstfeld and Rogoff 1995). In fact, the intertemporal approach brought about an analogy between the budget constraint and the external constraint that was only implicit in previous conventional analysis (Currie 1976).

Hence, in a world with developed financial markets a country may choose to smooth out spending patterns and delay the adjustment of the balance of payments for several periods. The conventional wisdom is that economies tend to be at the output level that corresponds to the natural rate of unemployment in the long run, and as a result in the long run the adjustment is done by variations of relative prices, either a deflation or a depreciation of the currency. In both cases, monetary policy is seen as the

main instrument to achieve the balance of payments equilibrium.

There are several limitations to this analysis. A crucial problem is the idea of a natural level of unemployment, which subtracts any relevance to changes in the level of activity in the balance of payments adjustment process. The natural rate of unemployment corresponds to the full employment level. It is worth noticing, however, that the natural rate has been conspicuous for its absence in the 1990s.

In the early 1990s in the US most macroeconomist agreed that the natural rate was around 6 per cent. If unemployment fell below that rate, then the economy would overheat and inflation would follow. Yet, by the mid-1990s unemployment had fallen to around 4 per cent and inflation was nowhere to be seen. The Federal Reserve Board was praised by the market for not hiking interest rates when unemployment rates started falling. Some economists argued then that the natural rate had fallen to around 4 per cent.

In fact, to understand the importance of adjustments in the level of activity one does not need to make a big effort. All balance of payments crises (e.g. the Latin American debt crisis of 1982, the Asian Crisis of 1997, Argentinian crisis of the early 2000s) were followed by severe contractions of the level of activity and increasing levels of unemployment. Further, as the contraction helps to reduce trade imbalances by reducing the level of imports, patterns of trade are also affected. Usually countries cut the imports of superfluous goods, and maintain imports of intermediary goods essential for production. These changes are seldom—if ever—determined by changes in relative prices.

In sum, a crucial element in the conventional view about balance of payments adjustment depends on a proposition that is—to say the least—difficult to defend in theory, and that has scant evidence in its support. Full employment or tendencies towards it are not a common feature of modern economies. So one could ask what would be the consequences of abandoning that assumption, and assuming a more pragmatic macroeconomic theory for the theory of balance of payments adjustment.

A second and interrelated critique of the dominant approach is the notion that the balance of payments is self-adjusting, and that led to themselves markets would adjust towards equilibrium. Intervention on the balance of payments is, however, pervasive, since markets have indeed a tendency to lead to balance of payments crises. Calvo and Reinhart (2000) have noted that even countries that claim to pursue flexible exchange rate policies tend to intervene in foreign exchange markets, showing signs of what they refer to as fear floating. The reason behind fear floating is the perception that the balance of payments will not adjust by itself.

The recent experience with balance of payments liberalization is a good illustration of this point. Some of the problems of the dominant view become clearer in light of a discussion of the causes of balance of payments crises.

### **Balance of Payments Crises**

The canonical model of balance of payments crises was developed by Krugman (1979), based on the work of Girton and Henderson (1976). According to this view the main cause of a balance of payments crisis is overspending. Governments tend to run



fiscal deficits, which in turn are financed by money emission, leading through the simple Quantity Theory of Money to inflationary pressures. The inflationary pressures imply that domestic goods become more expensive, leading to current account deficits (twin deficits) and, hence, to pressures for depreciation. Depreciation and the substitution effects that it provokes adjust the balance of payments, but a new crisis can only be avoided by fiscal adjustment.

Several authors extended the conventional story. In particular, it was noted that governments not only had to be fiscally responsible, but they had to be perceived to be fiscally responsible. In other words, credibility is the key to avoid balance of payments crisis, so creating a reputation for fiscal responsibility should be the main task of financial ministers around the world. In that case, it is not impossible to imagine a situation where a country suffers a balance of payments crisis even though it pursues market-friendly policies. Obstfeld (1986) shows, using a model that is in essence the same as the one developed by Krugman, that countries with pegged exchange rates are particularly vulnerable even if they pursue responsible fiscal and monetary policies.

To illustrate alternative views on balance of payments crises, one could concentrate on some specific episodes, and adequacy of the conventional approach. The German balance of payments crisis of the early 1920s, the collapse of Bretton Woods, the Latin American Debt Crisis of the 1980s, and the more recent Asian Crisis provide rich periods for analysis.

The conventional view that the balance of payments crisis was caused by fiscal irresponsibility was for example the interpretation of the balance of payments

crisis and the hyperinflation in Germany during the 1920s put forward by most economists (Bresciani-Turroni 1931). The German officials that had to deal with the day-to-day problems of running an economy under hyperinflationary conditions saw the problem, not surprisingly, from a different perspective. The most notorious defender of the so-called balance of payments theory was Karl Helfferich (Câmara and Vernengo, 2001).

For Helfferich the main cause of hyperinflation was to be found in the reparations of the Versailles Treaty. Helfferich argued that the permanent unfavourable trade balance, caused by the war and the impositions of Versailles, led to depreciation. This was the root of German problems. This idea that at the root of the balance of payments crisis is a terms of trade problem, or some other real cause that reduces the ability of the country to enter international markets in a competitive position would be taken again by Latin American structuralists (Prebisch 1959).

In this view, then, the conventional story is put upside down. A term of trade shock—or a reparations problem as in the case of Germany—imposes a financial burden on the balance of payments that cannot be paid out of the current account surpluses (when they exist). Hence, the country is forced to depreciate to generate the current account surpluses, leading to higher prices of imported goods and to inflation. Further, contraction of the domestic output level is also needed to cut imports to the minimum. Depreciation, one should add, in typical structuralist fashion, works by redistributing income to exporters—usually capitalists with higher propensities to save—and reinforcing the contractionary effects. Hence, the balance of payments is adjusted

by output contraction, and the income effects rather than the substitution effects are the one that count, as in the Keynesian interpretation of balance of payments adjustment.

The Latin American Debt crisis—that followed the Mexican default of August 1982—was also seen by conventional authors as the result of government overspending. Latin American neo-structuralists, on the other hand, saw it as the result of a combined terms of trade cum interest rate shock that made the foreign debt unserviceable, leading to depreciation and contraction (a lost decade was the nickname for the 1980s), as the only way to adjust the balance of payments. The similarity with the German balance of payments crisis of the 1920s, which also was associated to hyperinflation, was well noted by several Latin American authors (Câmara & Vernengo 2001).

The alternative view then would emphasize the role of terms of trade shock—prices of commodities, including oil fell considerably in the 1980s—and the interest rate shock caused by the hike of American rates by Paul Volcker, the then chairman of the Federal Reserve Board, as the two main causes for the debt crises. Faced with an increasing debt servicing bill and with reduced resources, Latin American countries were forced to contract the level of activity to reduce imports, and devalue their currencies—with both contractionary and inflationary effects—to adjust the balance of payments.

The collapse of the Bretton Woods system also provides an interesting contrast between the conventional and alternative interpretations of balance of payments crises. For the conventional neoclassical analysis the main cause of the demise of

Bretton Woods is associated to the inflationary pressures brought about by the expansionary fiscal policies in the U.S., and the propagation of these inflationary pressures through the international system. The increasingly expansionary fiscal policies of the 1960s—resulting both from the Vietnam War and the Great Society experiment of the Kennedy-Johnson administrations—led to growing balance of payments deficits. The U.S. deficits were initially considered instrumental for the working of the international monetary system that was desperately in need of dollars to obtain the essential imports of capital goods needed for reconstruction. However, by the late 1960s the accumulation of idle dollar balances started to put pressure on the money supply of the rest of world, leading to inflation. That is, according to the neoclassical logic, inflation was caused by the U.S. fiscal and monetary policies, and transmitted to the world as a result of the system of fixed parities.

The collapse of Bretton Woods is then related to the unwillingness of foreign countries to import U.S. inflation. That eventually broke the credibility of the fixed exchange rate commitments, and the willingness of the several central banks to cooperate in order to maintain the fixed parities. In other words, the Bretton Woods system failed because the fixed parity commitment was not credible in the face of accelerating inflation.

An alternative explanation for the inflationary pressures of the 1960s is possible though. This alternative explanation would minimize the effects of the U.S. expansionary fiscal policy in the demise of Bretton Woods. The Golden Age regime implied a commitment to full employment and the creation of a safety net for

unemployed workers. Additionally, the imposition of capital controls and the cheap money policies—which led to low real rates of interest—implied a favourable environment for workers. Parties with strong ties with the labour movement were in power in several Western countries, and this was tolerated, to a great extent, since it was considered a form of reducing the dangers of the Soviet menace. Further, full employment tends to increase the bargaining power of the working class.

In this environment, workers pressures for higher nominal wages would be usually accommodated. For a given real rate of interest, and a fixed nominal exchange rate, the only effect of rising wages would be higher prices. In sum, inflation was the result of wage pressures (cost-push) rather than the expansionary fiscal and monetary policies (demand-pull). In that sense, the abandonment of the fixed parities is not connected to the loss of credibility in the face of higher inflation, since the causes of inflation lay somewhere else (Vernengo 2004).

The alternative view emphasizes the role of financial liberalization in the collapse of the Bretton Woods regime. Davidson (1982) argues that the U.S. dollar represents the asset of ultimate redemption, and hence is used as the measure of international liquidity. As a result the U.S. benefits from a more liberal financial system, since the centrality of U.S. financial market allows it to attract funds to finance persistent current account deficits. For that reason, beginning in the 1960s the U.S. adopted a more self-centred financial policy, promoting financial openness in order to be able to face the growing current account deficits. It is the increasing financial openness of the 1960s, built into the American support for the

Euromarket that ultimately made the Bretton Woods system untenable.

Similar debates regarding causes of the Asian Crisis in 1997 resurfaced. The conventional view could not blame excessive fiscal spending as the main cause of the crises, since it was clear that most Asian countries had kept their budgets in surplus, as noted by Stiglitz (2002). On the other hand, the heavily interventionist policies that led to the so-called Asian Miracle—that is, the exceptional rates of growth of South East Asian economies—were to blame. The Asian values that have been considered essential for development were now seen as signs of crony capitalism. McKinnon and Pill (1998) for example argued that overregulation led to perverse incentives, and misallocation of resources. Hence, the process of development was illusory and financial markets would have to discount the actual costs of the oversized governments. In fact, Krugman (1994) in a widely acclaimed paper compared the Asian Tigers to the Soviet Union, and forecasted their collapse.

Several critiques of the conventional interpretation were raised, in which the role of financial liberalization and the lack of proper regulation on capital flows is seen as the main culprit for the balance of payments crisis. Chang (1998) argues that corruption and mismanagement only increased after the crisis leading to more lax regulation in the case of South Korea. The liberalization of the capital account of the balance of payments required for entry into the OECD group, and in general promoted by the IMF, is seen as the main cause for the balance of payments crisis.

Taylor (1998) provides a general overview of the effects of capital account liberalization around the developing world

during the 1990s, one in which the Keynesian view that international financial markets may very well be unstable and prone to crises is at the centre of stage. For example, a country may receive inflows of capital as a result of the increasing confidence of international financial markets on its ability to grow—say, for example, that the main reason is a recent history of fast growth. The inflows of capital lead to an increase of imports, and also to an appreciation of the domestic currency. Both effects tend to lead to a worsening of the trade account. The trade deficit is not necessarily bad. If these inflows were used to buy machines and equipment and lead to higher investment, and higher productivity, one would expect that exports in the future would more than compensate the initial deficits. In this case, the inflows would be sustainable and there is nothing to worry about.

On the other hand, if the inflows are used for consumption, and there are no prospects of higher exports, then one might be in trouble. Also, the appreciation of the currency, caused by the capital inflows, may force competitive firms out of business. This has hardly anything to do with lack of comparative advantage, since the only reason for failure is an appreciated exchange rate. Often countries in this situation would hike interest rates to attract capital flows and allow the trade deficits to continue for a while. This only makes things worse, since continuous inflows keep the exchange rate appreciated, and the high interest rates compress domestic investment. A trap of low growth and an unsustainable balance of payments are the results. The final crisis is usually triggered by some outside event that leads to capital flight and depreciation of the domestic currency.

Depreciation, however, is also contractionary, as we already saw. The depreciation means that those with debts in foreign currency (and usually revenues in domestic currency) have a prospect of a higher debt-servicing burden. Also, it is not uncommon to encounter that international debts were contracted short, while revenues are long. Hence, the currency and term mismatches mean that debtors are bankrupted, leading to lower levels of activity, lower growth and higher unemployment. It is the contraction of output, and the consequent fall in imports that helps adjust the balance of payments. Debt restructuring, unemployment and lower rates of growth for long periods tend to follow.

In other words, international financial markets can make the balance of payments adjustment process quite painful. That was the reason why Keynes advocated capital controls during the Bretton Woods conference. It is also behind the argument put forward by Stiglitz (2002) that the two countries that escaped the Asian Financial Crisis were China and India, the two countries with more comprehensive foreign exchange controls. Or more dramatically, this is also the reason why *The Economist*—a bastion for defenders of free trade in goods, services and capital flows for decades—has finally admitted that regarding liberalization of capital flows “the anti-globalists are on to something,” and that in this light “for some countries, imposing certain kinds of control on capital will be wiser than making no preparations at all” (Crook 2003). Even Krugman (1998) has advocated capital controls in the face of severe balance of payments crises.

The International Monetary Fund (IMF)—which was originally devised to support

countries with balance of payments problems—has also admitted recently that the experience with balance of payments liberalization has been less forthcoming than expected (Prasad et al 2003).

The recent negative perception regarding financial liberalization and the renewed defence of capital controls—as for example the discussion of Tobin taxes—is directly related to a resurgence of exchange-rate pessimism (Lane 2001; Obstfeld 2002; Obstfeld and Rogoff 1995). In recent years, many countries have suffered severe financial crises, producing a staggering toll on their economies, particularly in emerging markets. In Latin America there was a second Mexican crisis in December 1994, the so-called Tequila Crisis, the Brazilian crisis of January 1999 and the Argentinean meltdown of December 2001. Also, East Asia that survived the 1980s unscathed by financial crises and growing at miracle rates was hit by a crisis which spread from Thailand to other countries in the region during the second half of 1997, with contagion effects in Latin America as well as in Russia.

As noted by Pieper and Taylor (1998) the revival of the liberal creed has made balance of payments adjustment more painful for developing countries. The severity and frequency of the crises brought again the preoccupation with the negative effects of balance of payments disequilibria to long run growth to the centre of the debates.

### **Balance of Payments Constrained Growth**

The idea that the balance of payments constitutes a constraint to economic development can be traced back to Raúl Prebisch and other Latin American authors of the United Nations Economic Commission for Latin America (ECLAC).

Kaldor (1970) articulated a simple demand driven model of accumulation in which the balance of payments—foreign demand—had a crucial role. The development of his ideas dates back to the introduction of his technical progress function at the end of the fifties, and his interpretation of the slow rate of growth of Great Britain in the mid sixties, in which he developed the so-called Kaldor-Verdoorn Law, which states that productivity growth depends on demand stimulus.

Dixon and Thirlwall (1975:203) have correctly emphasized, “the main thrust of Kaldor’s argument is Hicks’s view that it is the growth of autonomous demand which governs the long run rate of growth of output.” In particular, the long run rate of growth is assumed to depend fundamentally on the growth of demand for exports. The growth of exports leads to higher rates of growth, which in turn force the system to generate innovations to keep pace with demand growth. The Kaldor-Verdoorn Law implies higher productivity and higher levels of income per capita, which generate a renewed increase in demand. Hence, a cumulative process of growth is put into motion.

Thirlwall (1979) showed that if a country is forced to keep its balance of payments in equilibrium, then the Kaldorian model implies that rates of growth should be proportional to the ratio of export growth to the income-elasticity of demand – also known as Thirlwall’s Law. In other words, if export growth is the main cause of GDP growth, and GDP growth leads to increasing imports, exports have to grow in tandem with imports to keep the balance of payments adjusted. McCombie and Thirlwall (1994) show that Thirlwall’s Law fits the data for most countries relatively well.

Dooley *et al.* (2003) also argue that export-led growth has been a staple source of growth in the periphery of the capitalist system. For them, the economic emergence of a fixed exchange rate periphery in Asia has re-established the United States as the centre country in a revived Bretton Woods international monetary system. They argue that the normal evolution of the international monetary system involves the emergence of a periphery for which the development strategy is export-led growth supported by undervalued exchange rates, capital controls and official capital outflows in the form of accumulation of reserve asset claims on the centre country. The success of this strategy in fostering economic growth allows some countries in the periphery to graduate to the centre.

If periods in which export-led growth are feasible in the periphery show that the balance of payments is central for development, it is also true that periods of financial liberalization show that the balance of payments is the main constraint.

### **Lessons and Policy Alternatives**

One important lesson from the last wave of balance of payments crises is that fundamentals do not seem to be the unique or even the main cause of them. Most countries in East Asia, for example, had fiscal surpluses, high private saving rates, and low inflation; and in most cases their exchange rates did not seem out of line. It should be then painfully obvious that the traditional measures recommended by the IMF—contraction and depreciation—will not solve the current set of problems (Stiglitz 2002).

It is not clear what will emerge from the general sense of inadequacy regarding the international financial architecture, but it is

clear that it will change. The Meltzer Report in the US wants to reduce the role of the IMF in the process of balance of payments adjustment to reduce the problems of moral hazard. That is, it is expected that free markets will impose discipline on economic agents. The crowds of people protesting against globalization in the World Bank-IMF annual meetings also want to eliminate the IMF, or at least their policies. Stanley Fisher, the ex-deputy director at the IMF recognized that the IMF would have to step in more frequently as a lender of last resort in order to make the balance of payments adjustment smoother. The world economy seems to be at a decisive juncture.

### **Selected References**

- Barro, R. and J-W. Lee (2002) "IMF Programs: Who Is Chosen and What Are the Effects?" *NBER Working Paper*, No 8951, May.
- Bresciani-Turroni, C. (1931) *The Economics of Inflation: A Study of Currency Depreciation in Post-War Germany*. London: Allen & Unwin, Ltd., 1937.
- Calvo, G. and C. Reinhart (2000) "Fear Floating", *NBER Working Paper*, No 7993, November.
- Câmara, A. and M. Vernengo (2001) "The German Balance of Payments School and the Latin American Neo-Structuralists", in L-P. Rochon and M. Vernengo (Editors), *Credit, Interest Rates and the Open Economy*. Cheltenham: Edward Elgar, 2001.
- Chang, H.J. (1998) "Korea: The Misunderstood Crisis", *World Development*, Volume 26, Number 8, pp. 1555-61.
- Crook, C. (2003) "A Cruel Sea of Capital: A Survey of Global Finance", *The Economist*, May 3.

- Currie, D. (1976) "Some Criticisms of the Monetary Analysis of the Balance of Payments Correction", *Economic Journal*, Volume 87, pp.771-73.
- Davidson, P. (1982) *International Money and the Real World*. London: Macmillan.
- Dixon, R. and A. Thirlwall (1975) "A Model of Regional Growth Rate Differences on Kaldorian Lines", *Oxford Economics Papers*, July.
- Dooley M.; D. Folkerts-Landau; and P. Garber (2003) "An Essay on the Revised Bretton Woods System", *NBER Working Paper*, No 9971, September.
- Eichengreen, B. (1996) *Globalizing Capital*. Princeton: Princeton University Press.
- Felipe, J. and M. Vernengo (2003) "Fallacies in Trade, Competitiveness and Macroeconomics", Asian Development Bank, *Policy Brief* No. 2003/10.
- Frenkel, J. and H. Johnson. (1976) "The Monetary Approach to the Balance of Payments", in J. Frenkel and H. Johnson (Editors), *The Monetary Approach to the Balance of Payments*. London: Allen & Unwin.
- Girton, L. and D. Henderson. (1976) "Central Bank Operations in Foreign and Domestic Assets Under Fixed and Flexible Exchange Rates", Federal Reserve Board International Finance. *Discussion Paper*, No 83, May.
- Harrod, R. (1933) *International Economics*. New York: Harcourt Brace.
- Hume, D. (1752) "Of the Balance of Trade", in R.N. Cooper, (Editor), *International Finance: Selected Readings*. Harmondsworth, Penguin, 1969.
- Kaldor, N. (1970) "The Case for Regional Policies", *Scottish Journal of Political Economy*, November.
- Krugman, P. (1979) "A Model of Balance-of-Payments Crises", *Journal of Money, Credit and Banking*, Volume 11, Number 3, Aug., pp. 311-325.
- Krugman, P. (1994) "The Myth of Asia's Miracle", *Foreign Affairs*, Nov/Dec, 73 (6) pp. 62-79.
- Krugman, P. (1998) "A letter to Malaysia's Prime Minister", *Fortune*, Sept 28, Volume 138, Number 6: 35-38.
- Krugman, P. and L. Taylor (1978) "Contractionary Effects of Devaluation", *Journal of International Economics*, Volume 8, pp. 445-56.
- Lane, P. (2001) "The New Open Economy Macroeconomics: A Survey", *Journal of International Economics*, August.
- McCombie, J. and A.P. Thirlwall (1994) *Economic Growth and the Balance of Payments Constraint*. New York: St. Martin's Press.
- McKinnon, R. and H. Pill (1998) "International Overborrowing: A Decomposition of Credit and Currency Risks", *World Development*, Volume 26, Number 7, pp. 1267-82.
- Meade, J. (1952) *The Balance of Payments*. Oxford: Oxford University Press.
- Obstfeld, M. (1986) "Rational and Self-Fulfilling Balance-of-Payments Crises", *American Economic Review*, Volume 76, Number 1, March, pp. 72-81.
- Obstfeld, M. (2002) "Exchange Rates and Adjustment: Perspectives from the New Open Economy Macroeconomics", *NBER Working Paper*, No 9118, August.
- Obstfeld, M. and K. Rogoff (1995) "Exchange Rate Dynamics Redux", *Journal of Political Economy*, Volume 103, Number 3, June, pp. 624-660.
- Pieper, U. and L. Taylor. (1998) "The Revival of the Liberal Creed: The IMF, The World Bank, and Inequality in a Globalized Economy", in D. Baker, G. Epstein, and R. Pollin (Editors),

- Globalization and Progressive Economic Policy*. Cambridge: Cambridge University Press.
- Prasad, E.; K. Rogoff; S-J. Wei; and A. Kose. (2003) *Effects of Financial Globalization on Developing Countries: Some Empirical Evidence*. Washington DC: IMF.
- Prebisch, R. (1959) “Commercial Policy in the Underdeveloped Countries”, *American Economic Review*, Volume 49, May.
- Sachs, J. (1981) “The Current Account and Macroeconomic Adjustment in the 1970s”, *Brookings Papers on Economic Activity*, Number 1, pp. 201-268.
- Stiglitz, J (2002) *Globalisation and Its Discontents*. New York: Norton.
- Taylor, L. (1998) “Capital Market Crises: Liberalisation, Fixed Exchange Rates and Market-Driven Destabilisation”, *Cambridge Journal of Economics*, Volume 22, Number 6, pp. 663-676.
- Taylor, M. P. (1990) “A History of Thought on the Balance of Payments”, in M. Ugur (Editor), *An Open Macroeconomics Reader*. New York: Routledge, 2002.
- Vernengo, M. (2004) “Bretton Woods”, in J. King (Editor), *The Elgar Companion to Post-Keynesian Economics*. Cheltenham: Edward Elgar.

*Matias Vernengo*  
*Department of Economics*  
*University of Utah*  
*matias.vernengo@economics.utah.edu*



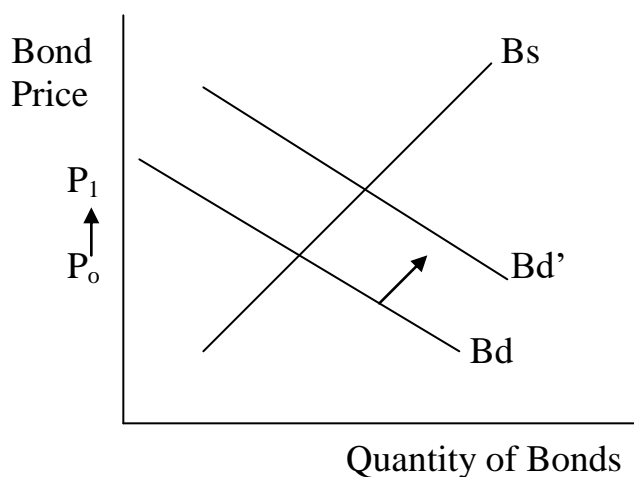
## Bond Market

*Caner Bakir*

### **Introduction**

A ‘bond’ is a tradable debt instrument. Since the market price of a bond is the discounted present value of a stream of future cash flows, bond prices move inversely to interest rates or yields. For example, if share prices rise, investor’s wealth increases, so savers would have more funds to save (i.e., the money supply increases). Accordingly, savers’ investment demand for bonds increases. Increase in money supply to purchase bonds causes interest rates to fall while pushing bond prices up (see Figure 1 below).

*Figure 1: Bond Price Movement*



Bonds offer a predictable long-term income stream, a low risk of default by the issuer, and prospects of capital gains as bond prices move in accordance with economic fundamentals (for a diagram of the bond market environment see Harwood 2000:8). The bond markets are composed of borrowers, lenders, regulators and intermediaries. They provide medium- to long-term financing to federal, state and local governments as well as corporations.

Governments issue bonds to finance budget deficits and to pay principal on maturing debt. However, objectives of the government bond market cannot be reduced to financing a government’s budget deficit. A large, liquid government bond market establishes a benchmark yield curve for pricing of corporate bonds. Further, such bonds, virtually free from credit risk, are used as collaterals for futures, options, and repo markets (secondary markets for bonds trading).

Apart from the government bond market, a corporate bond market also plays two significant roles in a national economy. First, economic growth and development cannot be generated without debt capital. Second, both bank lending and government public finance are not adequate to provide efficient long-term debt. Thus the development of a domestic debt market is vital to ensure a robust and efficient financial system as a whole.

The development of efficient and liquid bond markets are affected by three factors: first, “‘environmental factors’, such as the macroeconomic situation and the credit worthiness of the issuer; [second,] “‘institutional factors”, such as securities law, regulation, supervision, tax and accounting; [and third,] ‘market infrastructure’, including trading systems for cash and derivatives products, clearing and settlement systems, and central depository systems” (Sabatini 2001:300; see also ASEAN May 2002).

The size of the world bond market was US\$33 trillion in 2001 (Molinas and Bales 2002:1). As table 1 shows, governmental (i.e., government and quasi-government) and corporate sectors (i.e., corporations and agencies) had considerable share in the world bond market in 2001 with 36.2 per

cent and 29.7 per cent, respectively. In particular, agencies (e.g., mortgage originators) were the fastest growing sector due to strong growth in mortgage market. The share of central government's bonds in the world bond market, however, fell to 30 per cent in 2001 from 41 per cent in 1995. Continued budget surpluses in Scandinavia and English speaking countries (Australia and Canada) drove the decline of government bond markets although both in Japan and the EU markets continued to grow (see Molinas and Bales 2002:1-2).

*Table: 1 World Bond Market Sectoral Composition*  
(US\$, 1995-2001, % Share World Bond Market)

	2001	1999	1997	1995
Central Government	30.1	36.0	38.8	41.4
Quasi-Government	6.1	5.8	6.2	6.9
Agents	17.0	14.6	13.8	13.2
Corporate	29.7	28.7	26.9	27.1
Foreign	2.6	2.5	2.9	2.6
Eurobond	14.5	12.2	11.4	8.8

*Source:* Adapted from Molinas and Bales (2002:1)

Organization for Economic Cooperation and Development (OECD) countries have developed bond markets because “the focus on debt management and the upgrade of debt management capabilities, establishment interest rate, liquidity and currency benchmarks has helped to improve the transparency, predictability and liquidity of fixed income debt market. ... The increase in the prominence of the debt office has revolutionized, starting in the mid-1980s, how OECD governments finance themselves, and the critical role that fixed-income securities markets play in the financial sector. An effective management policy is always accompanied by a developed fixed-income market” (Blommestein 2001:186). In most countries, the development of bond markets helps “to

finance deficits, to sterilize large capital inflows, to make financial markets more complete by generating market interest rates that reflect the opportunity cost of funds at each maturity, to avoid concentrating intermediation uniquely on banks, and to the operation of monetary policy” (Turner 2002:2-3).

### **Bond Markets: Brief History**

As Neal (2003:1) puts it, “so many financial innovations throughout history, came from the exigencies of war finance.” In particular, early government bonds were mainly issued to finance the Eighty Years war expenditures, 1572-1648 (Neal 2003; Ferguson 1999). In referring to evidence from England, North and Weingast (1989) argued that the credibility of the monarch to repay the debt was a prerequisite for the emergence of an efficient public, and then private, debt market. Company bond finance had emerged in the 17<sup>th</sup> century to finance mainly overseas trade (Dickson 1967; Neal 1990, 2003). Dickson (1967:407) notes that “Company share-issues and bond-debts corresponded to, and were perhaps modelled on, the mortgage loans and personal bonds of private individuals.” Maturities of such bonds were as short as one month and up to three months (Dickson:408). The Dutch East India Company, which issued its medium-term corporate bonds on 26 October 1622 in Amsterdam, is regarded as the pioneer of the corporate bond finance (Neal 2003). Although Amsterdam “had enjoyed the fruits of early financial capitalism so spectacularly in its golden age of the seventeenth century,” capital markets emerged in London in the eighteenth century (Neal 2003:13). The level of economic development and state centralization in

England were two main factors behind the success of London.

There are four rival explanations of development of the corporate securities market (or capital markets where bonds and shares are issued by corporations for long-term finance) in banking and financial history literature: “First, historians generally hold the general level of economic development as the prime suspect of financial markets development...A second explanation points to the seeding role of prior public-debt markets...The third explanation stresses the role of rules favoring disclosure of financial information and curtailing insider trading on privileged information...The fourth and most recent...emphasizes the common law or civil law origins of the legal system...—common law is more apt than civil law to reduce contracting uncertainty between the parties to a security issue” (Verdier 2001:329-30).

Verdier (2001:330-31). emphasizes the role of domestic factors in the development of corporate securities markets in the nineteenth century: “markets developed as a result of conflict between corporate financiers and traditional sectors [i.e., agriculture, artisans, shopkeepers], mediated by politicians, and of which the outcome was influenced by the degree of centralization of state institutions...The development of a corporate market was part of a larger financial revolution involving the creation of money market and the concentration of banking.”

In the eighteenth and nineteenth centuries, investors did not have an access to fiscal and monetary data to judge default-risk, or inflation and currency depreciation. Ferguson (1999:21) argued that “nineteenth-century investors priced bonds in response

as much as to political news as to less accessible fiscal and monetary indicators.” According to Ferguson (p.18), “brokers and investors had to infer changes in fiscal and monetary policy largely from political events. Among the most influential bases for such inferences were the assumptions: 1. that a political move to the left, ranging from outright revolution to a change of ministry due to elections, would tend to loosen fiscal and monetary policy; 2. that a new and radical government would be more likely to pursue an aggressive foreign policy which might, in turn, lead to war; 3. that any war would disrupt trade and hence lower tax revenues for all governments; and 4. that direct involvement in war would increase a state’s expenditure as well as reducing its tax revenues, leading to substantial new borrowings.” Not surprisingly, the nineteenth-century bond market was heavily influenced by “information and assumptions about changes of international or domestic instability” (p.23).

In the twentieth century, before 1963, long-term capital in international markets was primarily raised through ‘foreign bonds’ issued by foreigners in another country, denominated in the currency of the host country and issued in accordance with its regulatory requirements. Then, Eurobonds—fixed-income securities denominated in various currencies that were issued by an international syndicate—became the principal form of financing in the bond markets with maturities of less than 10 years. Eurobonds are always available in bearer form, preserving anonymity of ownership. In contrast to the domestic bond markets, the Eurobond market is substantially a wholesale market where governments, international organizations (e.g., World Bank, European Investment

Bank), multinational corporations, and professional money managers are some of the market participants.

As Helleiner (1994:81) argued, “states have played a much more central role in the reemergence of global finance.” Specifically, Britain and the United States provided a support for the creation of Eurobond markets: “Britain provided a physical location for the market, permitting it to operate in London free of regulation. ...The British enthusiasm grew out of its historic commitment to maintain London’s position as an international financial center” (pp.82, 98). The British plan to establish London as a Eurobond center was successful “largely due to the plan’s coincidence with the United States (US) capital controls program which was intended to be a solution of the country’s growing balance-of-payments problems in the 1960s” (p. 84). The reasons for American support for the Eurobond market were twofold: “[First,] the New York financial community and multinational industrial firms demanded access to the Euromarket as compensation for the freedom they had lost as a result of the US capital controls program. [Second,] the United States found in the Euromarket an important tool for encouraging foreigners to help finance its external deficits” (pp. 99-100).

The global bond market has shown rapid growth between the 1990s and 2003. For example, this market measured in the USD billion equivalent to the nominal value of outstanding issues in all currencies increased from \$2.3 billion in 1995 to \$6.9 billion as at 30 June 2003. In the same period, the size of the Eurobond market in the global bond markets reached \$3.4 billion whereas that of the US bond market and the United Kingdom

(UK) was \$3.2 billion and \$543 million, respectively.

### **European Bond Markets**

Following the European Monetary Union (EMU) and the introduction of the euro on 1 January 1999 as a single currency, the European Union bond markets reached the size of the Japanese bond market and were second to the US-dollar market (Steinherr 2002). There are three key benefits for the European bond markets coming from the EMU and euro: first, “stable and low inflation should reduce the economic risk, driving down risk premia, and ultimately enabling investors to adopt longer time horizons for their investment”; second, the adoption of the euro “eliminates currency risk in cross-border investments”; third, “the replacement of national currencies by the euro should lead to the disappearance of a regulatory ‘home bias’, imposed on many institutional investors facing strict limits on the extent of currency mismatches that they are allowed to bear” (Steinherr:151). Indeed, the European bond market was positively affected by the euro. For example, the elimination of currency risk in the euro-zone removed market fragmentation and increased the size of government and corporate bond issues; the outstanding stocks of government bonds for Euro 11 (\$2.4 billion) exceeded those of the US (\$1.74 billion) and Japan (\$1.73 billion) in 2000 (see Steinherr:152).

Following the introduction of the euro, several structural factors have influenced the European bond market developments: “the consolidation trend in the banking [sector]; the drive to an efficient use of banks’ capital and balance sheets; increased merger and acquisition activity financed by corporate bond issues; and the integration of financial

markets in the EU area” (Hanfland 2001:195). Unsurprisingly, there has been remarkable growth in the non-government bond market. For example, “the share of the outstanding bonds of corporations has doubled to 16 per cent [between 1995-2000] [and] the share controlled by collateralized debt...has increased” (see Steinherr 2002:154). There have also been remarkable changes in the structure of borrowers in the euro capital markets: first, “Compared with the 8 per cent market share of non-government outstanding bonds in 1995, the corporate sector has captured about a fifth of the total issuance since the launch of the euro; second, the share of new bond issues originated by banks is substantially lower than in the past while the collateralized bond market has been growing sharply”; third, ‘supranationals have reduced their calls on the euro market while regional and local authorities have started to tap the market” (Steinherr 2002:155).

Institutional investors, such as insurance companies and pension funds, tend to hold long-term debt and play a key role in bond market development (Carmichael and Pomerleano 2002). Thus, in developed countries, corporate bond activities are dominated by activities in the primary market of corporate investors. In particular, financial institutions such as life insurance companies and superannuation funds issue corporate bonds (i.e., financial assets traded in the secondary markets) to hold as assets against their liabilities until the bonds mature. The corporate bond market in the US is larger than non-US developed countries (Endo 2001). For example, the absolute size of US-dollar denominated non-government bonds (\$24.8 billion) exceeds that of the euro (€16.3 billion) in 2000, the euro non-government market has undergone

significant changes. In part, this is because of a distinct feature of the US corporate bond market (see Steinherr 2002, Table 7.6). Before the removal of restrictions on interstate banking in 1994, state-based commercial banks could not meet the funding needs of companies that expanded their business across state boundaries; nor could they meet the long-term capital needs of such companies (Endo 2001:242). As a result, the corporate bond market along with the stock markets provided a mechanism through which to raise capital across the country as well as from abroad.

### **East Asian Financial Crisis**

One of the main reasons of the East Asian currency crisis of 1997 was a currency and maturity mismatch. As *Asiaweek* (19 December 1997) noted: “What Asia lacks, and Europe and the US have, is a deep, liquid and mature debt market where three things can occur. First, governments and corporations can borrow long to invest long, thus eliminating the maturity mismatch inherent in Asia. Second, corporations can issue paper in US-dollar, yen or euro currencies, with clearing and settlement in Asian times, thus eliminating currency mismatches and developing a truly deep Asian debt market along the lines of euro-dollar and euro-yen markets. Third, finance ministers in East Asian economies can foster a vibrant debt market with adequate risk management by investing their reserves in Asia.” One of the important lessons learnt from the 1997 East Asian currency crisis is that underdeveloped bond markets have encouraged an over dependence on domestic and foreign bank financing which creates risks and volatility in short-term capital movements. For example, “if issuance of bonds denominated in Asian currencies

increases, this would have the merit of enabling reduction in exchange risk for issuers and investors” (Kobayashi 2001:173). Capital intensive production technologies invested in the private sector for economic development, in a world of global finance, entail long-term risks. These risks can be better managed in a corporate bond market which mediates between long-term investments and long-term capital for private sector investments (Endo 2001:239).

Given the fact that bank lending in general is for the shorter term, lending through corporate bond markets rather than bank lending is important for developing economies. Banks are not well suited to finance long-term investments because their liabilities are largely short-term deposits. Maturity mismatches occur when short-term liabilities (e.g., deposits) are used to fund long-term assets (e.g., company loans). Further, structural neo-liberal reforms (e.g., privatization and increased role for the private sector in economic development) adopted by developing countries increased the role of a corporate bond market in financing investments in these countries. In specific terms, a corporate bond market is expected to play roles: “to diffuse stresses on the banking sector by diversifying credit risks across the economy, to supply long-term funds for long-term investment needs, to supply long-term investment products for long-term savings, to lower costs by avoiding a liquidity premium, to provide products the flexibility to meet the specific needs of investors and borrowers, and to reallocate capital more efficiently” (Endo p.238).

East Asian countries (Indonesia, Malaysia, Philippines, Singapore, Thailand, Republic of Korea, People’s Republic of China (PRC), Hong Kong, and Taipei)

traditionally rely on bank lending rather than bond markets. In bank-based financial systems (e.g., East Asian countries, emerging markets and Japan) majority of financial liabilities and assets consists of bank deposits and direct loans whereas in market-based financial systems (e.g., the UK and the USA) financial securities traded in the secondary markets constitute the bulk of financial assets (for a useful discussion and comparative analysis, see Zysman 1983; Vitols 2001).

Not surprisingly, Asian debt markets were underdeveloped as alternatives for bank lending in the 1980s and 1990s. For example, in the first quarter of 1999, bonds issued in Asian markets constituted less than 2 per cent of total bond issues of 766.7 billion US-dollars in global bond markets in the first quarter of 1999 (Sakakibara 2001:92). Apparently, Asian borrowers prefer to raise funds largely in Euro and New York markets. In the words of Sakakibara (p.92), the result “was the maturity and currency mismatch” in these markets. There have been three significant infrastructure problems making the major Asian bond markets (namely, Hong Kong, Singapore, Sydney, and Tokyo) unattractive: “First, the settlement and clearing system has been inadequate, requiring a much longer transaction time as compared to London, for example. Second, repo markets have not developed sufficiently. Third, although there are not direct regulations hampering transactions, such practices as withholding taxes, or the existence of multiple regulatory agencies has discouraged issuers. Also, in some countries the numbers of investors, particularly institutional investors in debts instruments, have been somewhat limited” (pp.91-2).

However, it should also be noted that East Asian countries have different levels of bond market development (see also above, note v). Hong Kong and Singapore belong to the first tier followed by Korea and Malaysia (the second tier), Philippines, Thailand, PRC and Taipei (the third tier). Indonesia has the least developed bond market among East Asian countries (ASEAN May 2002). Nevertheless, the development of liquid, sound and deep bond markets has become an obsession among policymakers within this region. Accordingly, the First and Second Round Tables on Capital Market Reforms in Asia were held in Tokyo in April 1999 and 2000, respectively (see OECD 2001). These meetings organized by the OECD and the Asian Development Bank Institute were international responses to bond market development in Asia. It was recognized that a large, efficient, liquid and well-regulated regional bond market can be achieved through well-regulated and efficient domestic bond markets. One of the suggestions to encourage market participants to start new market practices and to generate demand and interest for Asian bonds was to establish an Asian currency-based Asian Bond Index in the region. In a series of initiatives planned for 2003 in connection with the development of an Asian bond market, the Thai government is expected to issue baht-denominated bonds in Japan, while the Asian Development Bank is to issue a series of local currency bonds in India, Thailand, the Philippines, and Indonesia (Thai Press Reports 28 August 2003). Further, in August 2003 a group of 11 Asia-Pacific nations (Australia, China, Hong Kong, Indonesia, Japan, South Korea, Malaysia, New Zealand, the Philippines, Singapore and Thailand) established a US\$1 billion Asian Bond Fund 1 to purchase

dollar bonds issued by Asian governments or even corporate entities (Thai Press Reports 17 October 2003). As such, each Asian country reforms its own domestic bond market, harmonizes and standardizes regulatory frameworks and works for a regional forum for cooperation.

### **Bond Markets in Emerging Economies**

Emerging market international and domestic bonds outstanding amounted to 36 per cent of GDP in 2000, compared with only 24 per cent in 1994 (Turner 2002:1). In 2000, short-term debt constituted 31 per cent of total debt in these markets. In regard to the structure of domestic debt in emerging markets, 27 per cent of the total domestic debt was at a floating rate whilst 48 per cent was fixed. Inflation indexed and exchange rate linked debt constituted 16 per cent and 9 per cent of the total domestic debt in the same year (Turner, p.1). Institutional investors held a smaller proportion of this debt in emerging markets than was the case in the major industrialized countries. For example, assets of insurance companies were only 10.8 per cent of GDP whereas pension funds and pooled investment schemes did not exist in emerging markets, however, in G10 countries the figures were 62.4 per cent, 46.0 per cent, and 13.5 per cent for insurance companies, pension funds and pooled investment schemes, respectively (see Turner, p. 6). In regard to the structure of domestic debt markets in many emerging markets in Asia, Europe and Latin America, at the end of 2000, 48 per cent of the total domestic debt was fixed rate securities, 27 per cent was floating rate, 16 per cent was inflation-indexed and 9 per cent was linked to foreign exchange. These securities were short-term rather than long-term.

Emerging economies in Europe have developed domestic government bond markets (Batten, Fetherston, and Szilagyi 2004; Szilagyi, Batten, and Fetherston 2003). Large budget deficits are financed through the sale of government securities to public and private sector institutions in these countries. However, as Bonin and Wachtel (2003:32), in their analysis of bond markets in the transitional economies, put it, “there are only rarely active secondary markets for bonds[,] since most of the government debt is purchased and held by the banks.” Domestic corporate bond markets in these countries are tiny and a secondary market is virtually non-existent. As Bonin and Wachtel (2003:33), put it, “[o]ne of the reasons for this is that the financial institutions that can be expected to participate in these market–insurance companies, pension funds–are also underdeveloped.” Further, unstable macroeconomic environment evidenced in the high level of domestic interest rates, high and volatile inflation rates are regarded as the main impediments to bond market development in developing countries (IOSCO 2002).

It should be, however, noted that in the period between 1997-2001, there was a strong increase in the issuance of corporate bonds in emerging markets (IMF 2003:ch.3). For example, domestic corporate bond issuance increased to 31 per cent in 2000-01, from 5 per cent of total corporate domestic and international funding in 1997-99, whereas domestic bank credit fell from 52 per cent of total corporate funding to 40 per cent during the same period (Witherell 2003:9).

Also, the governments and enterprises in these countries have a chance to raise long-term capital in international bond markets

via sovereign borrowing in major currencies. There are four main reasons behind the slow development of private domestic debt markets in transition economies: “(1) the lack of transparency about corporate restructuring; (2) poor accounting and disclosure that make it impossible to monitor the use of funds and the absence of a legal framework to define clearly the rights of creditors; (3) power of insiders that gives creditors little influence over management and little recourse in the event of default; (4) bond pricing is difficult[,] since it requires some knowledge of default probabilities and expectations of asset recovery” (Bonin and Wachtel 2003:36; see also Blommestein 2001).

Bank intermediated finance is the single dominant source of industrial funds in emerging European countries. As Szilagyi, Batten, and Fetherston (2003:67) argue “the appropriate approach to meet future financing needs is through the development of viable domestic and international bond markets.”

Finally, a well-supervised and regulated banking sector is part of the infrastructure needed for the development of a sound bond market. In particular, interlocking directorates between banks and corporations through family–owned conglomerates create opaque insider relations where vast sum of money change hands without being exposed to tough market scrutiny in developing countries (Sharma 2001; Bonin & Wachtel 2003). Arguably, developed bond markets lead to a better allocation of capital.

### **Bond Markets and Governance**

As Eichengreen notes, “If markets are global, so must be their regulation, as must be the institutions through which that regulation takes place” (2002:53). The



Eurobond market, however, is virtually unregulated. In other words, banks or issuers are not required to “provide for investor protection, orderly markets, or courts of law in which to deal with disputes or abuses” (Smith and Walter 2003:54). The market is self regulatory: Eurobonds are listed on the London and Luxembourg stock exchanges and the International Securities Market Association (ISMA) (formerly the Association of International Bond Dealers formed in 1989) sets minimum trading requirements. The association is composed of institutions and dealers active in the Eurobond market and its head office is in Zürich. The association’s membership has exceeded 600 leading banks and investment houses in some 50 countries for at least 16 years. The principal task of the ISMA is the compilation of commercial usages for regulating Eurobond market transactions. Its mission is to introduce greater stability and order into the Eurobond market. The objectives of the ISMA are: to study and resolve technical problems affecting the international securities market; to implement and enforce rules applying the orderly functioning of the market as a self-regulatory body; to provide information, services and assistance to market participants; and to enhance relations between its members and between related domestic and international markets and its exchanges and regulatory agencies. Apart from the ISMA, the International Primary Markets Association (IPMA) was created in London in 1984. Its aim is to promote the interests of the lead managers of new issues of debt instruments in the international capital markets. The three main objectives of the IPMA are: to maintain high ethical standards among the lead managers; to

improve documentation; and to create closer relations with supervisory bodies.

For Coleman (1996), as financial markets become transnational, there is a case for self-regulation. Regulation of the conduct of business in Eurobond markets by the ISMA and the IPMA is an example of private Governance. However, there is, in Moran’s (1984) word, an “esoteric” (private, informal and technical) tendency in this private governance. Not surprisingly, self regulatory culture in Eurobond markets emphasizes informality and putting trust in the market players. The collapse of Barings, England’s one of most prestigious investment banks at the time, in July 1995 was due to the scale of exposures built up in derivatives markets by a single trader in the Singapore office. As Moran (2001:422) notes, following this ‘fiasco’, “one of the distinctive features of the regulatory state” in financial services industry came into prominence in the Britain; “the declining willingness to trust social actors to comply with rules, and a corresponding resort to more formal and open scrutiny, inspection and audit.” A similar cultural transformation at the global level, however, did not take place in spite of corporate collapses.

The post-war financial system was under the management of domestic public and intergovernmental international organizations. However, since the collapse of the system in 1971, Globalization of financial markets was marked by accelerating capital mobility and speculative financial transactions which undermined the roles of public authorities *vis-a-vis* private market actors (Helleiner 1994). Today, “[c]rucial decisions about the structure of international capital markets have often been taken in narrow transnational policy communities where private interests risk domination, to

the exclusion of broad segments of society and outside the traditional legislative process. In the absence of strong public authority over private market power, the global financial order is becoming incompatible with the imperatives of domestic political legitimacy” (Underhill & Zhang 2001:4; 2003:4).

Bond markets not only enable greater diversification of funding and investment channels but also strengthen corporate governance: “Compared with bank loans, fund-recycling through bond markets means spreading risks from the banking sector to a wider sector of the public. This, therefore, requires more transparency in the financial conditions of issuers, which is an urgent element of restructuring of the corporate sector” (Kobayashi 2001:173). Between 2000 and 2003, the global financial system witnessed a number of major shocks: the bursting of the equity market bubble; the Argentinean crisis and default; September 11; corporate scandals and bankruptcies. These shocks had an impact on bond markets; bond yields on government debt reached their lowest levels in 40 years whereas bond prices soared as investors became risk-averse, increasing holdings of government bonds while reducing lower-rated corporate bonds (Witherell 2003). In particular, corporate scandals and bankruptcies had a negative impact on corporate bond markets: “Reflecting the general deterioration in credit quality, corporate bond defaults [in 2002] amounted to some 3 per cent of bonds outstanding, well above the previous record of the early 1990s. Through the middle of 2002, credit-rating downgrades were at extremely high levels... Not surprisingly, investor appetite for lesser rated paper dropped sharply” (Witherell 2003:6). In particular, net issues

of non-financial corporate bonds declined by 60 per cent in the US whereas corporate bond issuance declined by 11 per cent in 2002 against the previous year. The financial sector bond issues in the US actually increased by 17 per cent in 2002 whereas same figure for Europe declined by some 44 per cent (see Witherell 2003). There was also a sharp reduction in the volume of non-government (i.e., financial and non-financial corporate sectors) new issues due to investor resistance to lower rated bonds and weak business investment (see Witherell, Figures 4 and 5). Further, the worldwide economic downturn during this period was also coupled with declining inflation and interest rates across developed nations. Issuance of government bonds accelerated in 2002 due to the weaker fiscal positions in some of the developed countries.

Governments have a vital role to play in the development of a sound bond market by providing the infrastructure required. This vital role can be performed through regulation and supervision to develop trust in financial markets (e.g., high disclosure standards to promote transparency and accountability) and to ensure compliance by participants with the market rules of conduct. In this context, global initiatives to influence national practices are considered: “the IMF and the World bank, and the Financial Stability Forum, with impetus from G7 governments, have launched a multi-prolonged effort to encourage industrial and developing countries to upgrade their financial practices and institutions. The focus of this effort is institutional arrangements in areas like data dissemination; fiscal, monetary, and financial policy transparency; banking regulation and supervision; securities and insurance regulation’ accounting; auditing;

bankruptcy; and corporate governance” (Eichengreen 2002:53). Providing policy prescriptions, however, are easier than their implementation. Then the fundamental issue is “how to develop institutions with the capacity to determine appropriate policies, implement them and stick to them until circumstances change” (p.59). Enhancing state’s what Weiss (1998, 2003) called ‘transformative capacity’ to manage change and to steer institutional adjustment in a world of global finance may be a credible answer: the negotiation process between the relevant state actors (which must be sufficiently insulated against sectoral interest groups) and organizations representing regulated sectors (which must have membership coherence to represent the whole of the sector) must be institutionalized and regularized, rather than ad hoc in character.

### Internet Sites

Bond Market Books.

<http://web.uflib.ufl.edu/cm/business/books/bondbks.htm>

RGE Monitor. [www.rgemonitor.com](http://www.rgemonitor.com)

### Selected References

- ASEAN. (2002) *Bond Market Development in East Asia: Issues and Challenges*. Regional Development Unit, Manila: Asean Development Bank.
- Asia Week. (1997) “The Asian Debt Market” 19 December.
- Batten, Jonathan, Thomas A. Featherston, and Peter Gabor Szilagyi. (2004) (Editors) *European Fixed Income Markets : Money, Bond and Interest Rate Derivatives*. London & New York: Wiley.
- Blommestein, Hendrikus. (2001) “Key Policy Issues in Fixed Income Securities Markets in Emerging Market Economies”, *Bond Market Development in Asia*. Paris: OECD, pp. 179-201.
- Bonin, John and Paul Wachtel. (2003) “Financial Sector Development in Transition Economies: Lessons from the First Decade”, *Financial Markets, Institutions and Instruments*, Volume 12, Number 1, pp.1-66.
- Carmichael, Jeffrey and Michael Pomerleano. (2002) *The Development and Regulation of Non-Bank Financial Institutions*. Washington: World Bank.
- Coleman, W. D. (1996) *Financial Services, Globalisation and Domestic Policy Change*. New York: St. Martin's Press.
- Dickson, P. G. M. (1967) *The Financial Revolution in England*. London, Macmillan.
- Endo, Tadashi. (2001) “Corporate Bond Markets Development”, in *Bond Market Development in Asia*. Paris: OECD, pp. 237-96.
- Eichengreen, Barry. (2002) “Capitalizing on Globalization”, *Asian Development Review*, Volume 19, Number 1, pp.14-66.
- Ferguson, Niall. (1999) *Wars, Revolution and the International Bond Market from the Napoleonic Wars to the First World War*. Paper presented at the Yale International Centre for Finance, 15 October, Yale University.
- Harwood, Alison. (2000) (Editor) *Building Local Bond Markets An Asian Perspective*. Washington DC: World Bank.
- Helleiner, Eric. (1994) *States and the Reemergence of Global Finance: From Bretton Woods to the 1990s*. Ithaca, NY: Cornell University Press.
- IMF. (2003) *Global Financial Stability Report, Market Developments and Issues*. Washington DC: IMF.

- IOSCO. (2002) *The Development of Corporate Bond Markets in Emerging Market Countries*.
- Kobayashi, Toshio. (2001) "Development of International Bond Markets in the Region—New Miyazawa Initiative" in *Bond Market Development in Asia*. Paris: OECD, pp. 173-75.
- Moran, M. (1984) *The Politics of Banking: the Strange Case of Competition and Credit Control*. London: Macmillan.
- Moran, M. (2001) "Not Steering but Drowning: Policy Catastrophes and the Regulatory State", *The Political Quarterly* Volume 17, Number 14, pp.414-427.
- Molinas, Cesar and Gioia P. Bales. (April 2002) *Size and Structure of the World Bond Market*. London: Merrill Lynch.
- OECD. (2001) *Bond Market Development in Asia*. Paris: OECD.
- Neal, Larry. (1990). *The Rise of Financial Capitalism: International Capital Markets in the Age of Reason*. Cambridge: Cambridge University Press.
- Neal, Larry. (2003) *Venture Shares of the Dutch East India Company*. Paper presented at Yale School of Management Conference, 6-7 March, Yale University.
- Sabatini, Giovanni. (2001) "The Role of Regulators in the Development of Government Securities Markets: An Overview", *Bond Market Development in Asia*. Paris: OECD, pp. 299-318.
- Sakakibara, Eisuke. (2001) "A Brief Overview: Creation of Asian Bond Market", *Bond Market Development in Asia*. Paris: OECD, pp. 91-4.
- Sharma, Krishnan. (2001) "The Underlying Constraints on Corporate Bond Market in Southeast Asia", *World Development*, Volume 29, Number 8, pp. 1405-19.
- Smith, Roy C. and Indigo Walter. (2003) *Global Banking*. Oxford: Oxford University Press.
- Steinherr, Alfred. (2002) "Europe in the Global Market Place", in Henryk Kierzkowski (Editor), *Europe and Globalization*. Basingstoke: Palgrave.
- Szilagyi, Peter, Jonathan Batten, and Thomas Fetherston. (2003) "Disintermediation and the Development of Bond Markets in Emerging Europe", *International Journal of the Economics of Business*, Volume 10, Number 1, pp. 67-82.
- Turner, Philip. (2002) "Bond Markets in Emerging Economies: An Overview of Policy Issues", *The Development of Bond Markets in Emerging Economies*. BIS Papers, No. 11.
- Underhill, Geoffrey and Zhang Xiaoke. (2001) *Global Structures and National Imperatives: In Search of Normative Underpinnings for International Financial Order*. A paper presented at From Naples to Genoa: A New World of Finance and Development Conference, Rome.
- Underhill, Geoffrey and Zhang Xiaoke. (2003) (Editors) *International Financial Governance under Stress: Global Structures versus National Imperatives*. Cambridge: Cambridge University Press.
- Verdier, Daniel. (2001) "Capital Mobility and the Origins of Stock Markets", *International Organization*, Volume 55, Number 2, pp.327-56.
- Vitols, Sigurt. (2001) *The Origins of Bank-Based and Market-Based Financial Systems: Germany, Japan, and the United States*. Wissenschaftszentrum Berlin für Sozialforschung. <http://skylla.wz-berlin.de/pdf/2001/i01-302.pdf>
- Weiss, Linda. (1998) *The Myth of the Powerless State*. Ithaca, NY: Cornell University Press.

- Weiss, Linda. (2003) *States in the Global Economy: Bringing Domestic Institutions Back in*. Cambridge: Cambridge University Press.
- Williams, Stephen. (2001) “Transparency in the US Debt Market”, *Bond Market Development in Asia*. Paris: OECD, pp. 319-33.
- Witherell, William. (2003) *Global and Local Bond Markets: A Financial Stability Perspective*. A paper delivered at the Fifth Annual OECD-World Bank Bond Market Forum, Washington.
- Zysman, J. (1983) *Governments, Markets, and Growth Financial Systems and the Politics of Industrial Change*. Ithaca, NY: Cornell University Press.

*Caner Bakir*  
*Koç University*  
*Istanbul, Turkey*  
*cbakir@ku.edu.tr*

## Budget Deficits

*Malcolm Sawyer*

### **Introduction**

Budget deficits, resulting from government expenditure exceeding taxation and resulting in the growth of public debt have been around for a long time as illustrated by the following quote. ‘The national budget must be balanced. The public debt must be reduced; the arrogance of the authorities must be moderated and controlled. Payments to foreign governments must be reduced, if the nation doesn't want to go bankrupt. People must again learn to work, instead of living on public assistance.’ (Cicero:106-43 BCE). This serves to illustrate a long-standing view on budget deficits that they are a result of the profligacy of governments and that they cast burdens on future taxpayers, and that measures were necessary to force governments to restrain their spending below their revenues. The Gramm-Rudman Act of 1985 in the USA sought to impose a balanced budget over a five year time horizon. The Stability and Growth Pact governing the macroeconomic policies of national governments within the eurozone, has the aim of a maximum budget deficit equivalent to 3 per cent of GDP and budget in balance or small surplus over the business cycle (and discussed further below). These represent attempts (both of which have failed) to impose ‘financial discipline’ on governments.

In general, upto and including the Second World War, budget deficits were largely associated with war (including civil war). The United States provides an example of this. ‘The pattern of wartime debt expansion can be seen in the War of 1812 when the national debt nearly tripled between 1811

and 1816, in the Mexican War era when the debt more than quadrupled between 1845 and 1851, in the Civil War when the debt increased forty-two-fold between 1860 and 1866, in the Spanish-American War era when the debt rose 50 percent between 1893 and 1899 (although the larger part of this increase occurred before the war), and in World War I when the debt increased twenty-one-fold between 1914 and 1919. World War II also fit the pattern: the debt increased nearly sixfold between 1939 and 1946.’ In the post-war period, budget deficits have become a common feature of most industrialised economies: for example using the OECD *Economic Outlook* data set over the period since 1960 of the 22 countries reviewed 17 had on average run budget deficits and 5 surpluses with an average (over all countries) budget deficit of around 1 ¾ per cent of GDP.

### **Definitions**

We begin by providing some definitions concerning budget deficits. The total budget position (BP) is defined as

$$(1) \quad BP = G - T$$

being the difference between government expenditure (G) and tax revenue (T). Where the budget position is positive it is a surplus budget, and where negative a budget deficit. Government expenditure here includes both current and capital expenditure and interest payments on debt. The precise definition of the budget deficit is not always straightforward in practice as a result of disagreements over what should be included in the public sector and in government expenditure. For instance, should expenditure by a non-government body whose activities are formally or informally



underwritten by government be included as “government spending”? Also, what should count as tax revenue? Should, for instance, payments to government which are ‘voluntary’, such as motor vehicle license fees, be included as taxation?

The primary budget deficit (hereafter  $D'$ ) excludes interest payments from the calculation. The deficit is usually calculated in nominal terms, taking no account of inflation. However, the rate of interest paid on government bonds (hereafter  $B$ ) is likely to be higher when inflation is higher, but the real value of those bonds is declining and the holders of the bonds are suffering capital losses in real terms and the government gaining from the decline in capital value. The budget deficit could then be calculated in real terms to make allowance for these changes in real capital values.

Government expenditure can be divided into current expenditure and capital expenditure. A current budget deficit is the difference between current expenditure and tax revenue. Capital expenditure is defined in terms of expenditure on net capital formation and the depreciation on existing capital included in current expenditure.

An unbalanced budget, whether deficit or surplus, leads to changes in the liabilities of the government. There is a basic funding identity for the budget deficit, namely:

$$(2) \quad G - T = DM + DB$$

where  $DM$  is the change in the amount of Central Bank money, sometimes referred to as base or outside money and corresponding to the monetary aggregate  $M0$  in the usual terminology and  $DB$  are changes in the stock of interest bearing bonds.

One interpretation of this equation is to focus on the role of changes in base money,

and to argue that government deficits are financed by governments ‘printing money’, a view which is often linked with a monetarist approach to inflation viewing budget deficits leading to increase in money which in turn leads to inflation. But a Keynesian interpretation is that budget deficits are funded largely by government borrowing. In general governments do not directly ‘print’ money but rather borrow (sell bonds) to provide for the difference between expenditure and revenue. When expenditure is less than revenue, government is retiring bonds. The Central Bank performs open market operations through which the monetary base,  $M0$ , is provided in exchange for bonds from the private sector. The resulting change in  $M0$  is determined by the desires of the private sector (and specifically their requirements for increases in money as the economy expands). At a consolidated balance sheet in which the Central bank is included with general government the above equation then holds.

When the budget is in deficit, clearly the amount of bonds and hence interest payments will be rising, and this raises the question of the sustainability of a budget deficit: that is, can a deficit continue forever? We return to this question below. When the budget is in surplus, government debt is being retired, and a similar question of sustainability arises. A continuing surplus would eventually lead to no government bonds and no base money, after which a government running a surplus would be buying up private sector assets.

### **National Account Identities**

Many of the debates over the use of budget deficits can be understood by reference to a well known national accounts identity:

$$(3) S - I = X + NY - M + G - T$$

The left hand side is net private savings, that is private domestic savings,  $S$ , minus private investment (gross domestic fixed capital formation). The first term on the right hand side is the current account on foreign transactions: being the difference between exports,  $X$ , net income inflow,  $NY$ , and imports,  $M$ ; and the current account is the negative of the capital account.  $G-T$  is the budget deficit.

The first (and very influential) interpretation of this identity can be readily seen from a slight rewriting of the equation as:

$$(4) S = I + X + NY - M + G - T.$$

It then appears that if the budget deficit rises, investment or the current account position have to decline, and the emphasis has usually been placed on the decline in investment. This is generally referred to as ‘crowding out’, and the mechanism is often portrayed as budget deficit means borrowing and as the government seeks to borrow this puts upward pressure on interest rates. Investment is assumed to be sensitive to the rate of interest and hence declines. But it may be noted that there is an implicit assumption made here, namely that there is a given volume of savings which has to be ‘shared’ out between investment, budget deficit and lending abroad (as the other side of the current account position).

The second interpretation of equation (2) is the Kaleckian-Keynesian one. It views savings (and indeed other elements in the equation) as depending on the level of income, and in turn the level of income depending on the level of aggregate demand. Thus an increase in investment expenditure,

in the current account surplus or in the budget deficit, would generate increases in income and thereby increases in savings. Further, by reference to equation (3) it can be said that an excess of savings over investment can only be realised if there is some combination of current account surplus and budget deficit.

The case for the use of fiscal policy and hence for governments to operate with an unbalanced budget (whether in surplus or deficit) arises from the simple Kaleckian-Keynesian proposition that the level of private aggregate demand does not generally correspond to a high level of economic activity. Further, there is no automatic market mechanism to bring aggregate demand to an appropriate level compatible with the desired level of economic activity. The notion that the budget should always be in balance (or even on average in balance) is rejected on the grounds that a balanced budget is generally not compatible with the achievement of high levels of aggregate demand. From this perspective, the budget deficit is to be used to mop up ‘excess’ private savings (over investment), and the counterpart budget surplus used when investment expenditure exceeds savings (at the desired level of economic activity). It follows, though, that a budget deficit is not required when there is a high level of private aggregate demand such that investment equals savings at a high level of economic activity (and a surplus would be required when investment exceeds savings at the desired level of economic activity). This can be expressed as the government should set tax and expenditure such that the resulting budget deficit is given by :

$$(2a) G - T = S(Y_f) - I(Y_f) + Q(Y_f) - X(WY)$$



where  $Y_f$  is the intended level of income (which may be thought of as equivalent to full employment or to some supply side constraint),  $WY$  is world income (which is taken as given for the purposes of this equation). A tendency for savings to run ahead of investment leads to the view that a budget deficit is required (in the absence of any tendency for balance of trade surplus). But it is a short-fall of investment over savings that creates the requirement for a budget deficit: in the absence of any such short fall (in *ex ante* terms) there is no need for a budget deficit. The analysis of budget deficits should then be undertaken in a context, which at least allows for the emergence of an excess of (*ex ante*) savings over (*ex ante*) investment corresponding to high levels of income.

Equation (2a) can readily be re-written as:

$$(2b) \quad G - T = S - I + M - (X + NY)$$

and in this form has been interpreted in terms of the so-called twin deficits problem, that is budget deficit and deficit on the current account. As an accounting proposition, if net private savings ( $S - I$ ) is relatively small, then a budget deficit and current account deficit will change together. In a similar vein, if  $S - I$  is relatively stable, then an increase in the budget deficit (left hand side) would be associated with an increase in the current account deficit (second term on the right hand side). The coincident emergence of budget deficit and current account deficit in the USA in the 1970s and 1980s lead to much discussion of the twin deficits issue. First, note that a current account deficit raises the same question of sustainability as a budget deficit does since a current account deficit involves

the build up of foreign debt (or running down of overseas assets). However, some interpreted the twin deficits in a causal sense, arguing that the budget deficit was causing the current account deficit. But, of course, there may well be a causal relationship in the other direction, namely that the current account deficit ‘causes’ the budget deficit ; or there may be no causal links and it arises from the accounting arithmetic.

Governments draw up expenditure plans and set tax rates, but do not directly set the budget deficit which depends on private spending decisions. Expenditure plans may not come to fruition (e.g. hold ups in expenditure programmes) and some expenditure (notably on unemployment and related benefits) depends on the level of economic activity. Tax revenues depend on tax rates set and the level and structure of economic activity. Hence it can be said that the budget deficit is endogenously determined by the interaction of the activities of the private sector and the tax and expenditure plans of the government.

### **Perspectives on Budget Deficits**

This consideration of different perspectives on budget deficits starts with what may be seen as the orthodox view, often termed the ‘Treasury view’, ‘whereby loan-financed expenditures were regarded as diversionary or inflationary and thus provided no net addition to employment’ (Middleton 1985:145). Howson (1975:92) indicates that “the standard argument against unbalancing the budget now ran [in the early 1930s] [in terms that]... it would undermine confidence which was necessary for the cheap money policy. ... [I]t was unnecessary because deficits would bring about inflation only insofar

as they were financed by creating credit, something which could be done better (i.e. without the deleterious effects on confidence and without increasing the National Debt) by a cheap money policy.”

The term ‘Treasury view’ was used to portray the views of the British Treasury in the early-mid 1930s prior to the Keynesian revolution. There have, though, been debates as to how far this was their views and the nature of the mechanisms which would block any positive effects from budget deficits (e.g., see Middleton 1985:Ch 8).

One particular part of the ‘Treasury view’ which has largely disappeared is any notion that attempts should be made to balance the budget on an annual basis. It is generally recognised that the budget deficit outturn depends on what happens to the levels of income and expenditure in the economy during the course of the year. If economic activity is relatively low, then tax receipts fall (and in a progressive tax system fall proportionately more than the fall in income), and some elements of government expenditure (notably payments on unemployment benefits) rise. Budget deficits are then sensitive to the business cycle. For example, the European Commission has estimated that the sensitivity of the budget balance to output is around 0.5 per cent for the EU, that is a 1 per cent fall in GDP will increase the budget deficit by 0.5 per cent (Buti et al. 1997:7).

The budget position then acts as a form of automatic stabiliser. When there is an economic downturn, the tendency of tax revenues to fall and some parts of government expenditure to rise helps to soften the extent of the economic downturn. The other side of that coin obviously is that in economic upturn, rises in tax revenue

serve to damp down the degree of upturn. Attempts to rectify the budget deficit could worsen the situation: when demand and economic activity is low, raising taxes or cutting government expenditure would reduce aggregate demand and thereby economic activity further.

The Stability and Growth Pact (SGP) which has been put in place as part of the operation of the European single currency can be seen as linked with this Treasury view. In practice the SGP has not been fully adhered to and a number of national governments have run budget deficits which exceed the limits imposed by the SGP. A major feature of the SGP is the requirement that the national budget deficit does not exceed 3 per cent of GDP, and failure to meet that requirement could lead to a series of fines depending on the degree to which the deficit exceeds 3 per cent. It is also necessary for national budgetary policies to “support stability-oriented monetary policies. Adherence to the objective of sound budgetary positions close to balance or in surplus will allow all Member States to deal with normal cyclical fluctuations while keeping the government deficit within the reference value of 3% of GDP”. Furthermore, “Member States commit themselves to respect the medium-term budgetary objective of positions close to balance or in surplus set out in their stability of convergence programmes and to take the corrective budgetary action they deem necessary to meet the objectives of their stability or convergence programmes, whenever they have information indicating actual or expected significant divergence from those objectives” (Resolution of the European Council on the Stability and Growth Pact, Amsterdam 17 June 1997).

A related approach but which focuses on the distinction between current expenditure and capital expenditure is now generally described as the ‘golden rule’. This was adopted by the incoming Labour government in the UK in 1997, with a ‘Code of Fiscal Stability’. This was “the golden rule—on average over the economic cycle, the Government will borrow only to invest and not to fund current spending; and, the sustainable investment rule—the public sector net debt as a proportion of GDP will be held over the economic cycle at a stable and prudent level” (H M Treasury 1999:19).

Maintaining the public sector net debt ‘at a stable and prudent level’ relative to GDP has been interpreted as ‘Other things being equal, net debt will be maintained below 40 per cent of GDP over the economic cycle’ (H M Treasury 2005:17). It can also be noted that debt which arises from the operation of the Private Finance Initiative (PFI). Under the PFI, a private company generally constructs and finances an investment project (say building a school or hospital) which it then leases back to the government and typically provides a range of services related to that project (e.g. providing maintenance). Hence calculation of debt and deficit are not unambiguous.

The appeal of the ‘golden rule’ arises from the notion that borrowing is not undertaken for current expenditure but is undertaken for capital expenditure, perhaps under the misapprehension that capital expenditure by government generates future income (for government), which can be used to pay off the borrowing. But capital expenditure by general government (which is here distinguished from capital expenditure by public corporations) does not yield any direct future ‘profit’ to the government. Some forms of capital

expenditure (e.g. roads) may aid national prosperity and thereby raise national income and tax revenue, but some forms clearly do not (e.g. military equipment). But many forms of current expenditure also aid future growth, most notably that on education, health services.

Keynes (1980) appeared to have argued for some form of ‘golden rule’ when he argued for public investment to be set such that Private Investment plus Public Investment equals Savings, and advocated that “in peace-time budgets through the Chancellor making a forecast of capital expenditure under all heads, and comparing this with prospective savings, so as to show that the general prospective set-up is reasonably in accordance with the requirement of equilibrium. The capital budget will be a necessary ingredient in this exposition of the prospects of investment under all heads. If, as may be the case, something like two-thirds or three-quarters of total investment will be under public or semi-public auspices, the amount of capital expenditure contemplated by the authorities will be the essential balancing factor. This is a very major change in the presentation of our affairs and one which I greatly hope we shall adopt. It has nothing whatever to do with deficit financing” (p.352). However, there is a crucial difference, namely that the level of public investment should be set at such a level that it takes up the difference between the savings forthcoming at full employment and private investment. In terms of equation 2 above (and ignoring the foreign sector component) the budget deficit on the left hand side would be equivalent to the level of public investment and pitched at a level equal to  $S(Y_f) - I(Y_f)$ .

There is a different, and we would argue more appropriate, way of drawing the

distinction between current and capital expenditure. That distinction is between those forms of expenditure the effects of which are short lived and those forms of expenditure whose effects are long lasting. Thus, expenditure on education would (nearly) all be regarded as long lasting in that the education provided has an impact on the individual receiving the education for many years after. Then virtually all education expenditure would be treated as capital expenditure since the benefits of education now will accrue over many years into the future.

Capital expenditure from the perspective of adding to economic growth and future taxation may be quite different from what should be seen as capital expenditure from the perspective of the acquisition by the government of capital assets. Under the former perspective, much of the expenditure on education, training, research and development, some parts of health care, road construction would be treated as capital expenditure.

The Stability and Growth Pact of the eurozone is a clear example of seeking to operate fiscal policy on the basis of arithmetic rules on the budget deficit. The underlying view is that the economy is inherently stable with strong forces leading it to full employment. The role of fiscal policy is to stand back and let 'market forces' operate. A major argument used against the use of fiscal policy is that there are adjustment mechanisms which do ensure such high levels of demand. In this regard, the major theoretical argument has been the operation of the real balance effect (Pigou effect): low demand generates falling prices and rising real value of the money stock and wealth, which stimulates demand. The level

of aggregate demand is (eventually) brought into line with the supply-side equilibrium.

But it is well known (at least since Kalecki 1944a) that the real balance effect relies on 'external' money with net worth to the private sector and to the stock of money remaining unchanged in the face of price changes. In a world of largely bank credit money, the amount of 'external' money is relatively small: for example in the UK the ratio of M0 to GDP is less than 4 per cent; a price fall of 10 per cent would increase real value of M0 by the equivalent of 0.4 per cent. With a wealth effect on consumption of the order of 0.02 to 0.05 (OECD 2000:192), aggregate demand would change by the order of 0.01 per cent (for a decline of 10 per cent in the price level). As prices fall, the demand for M0 would fall and hence the stock of M0 would also fall. The empirical relevance of the real balance effect has long been doubted. Further, there are good reasons to doubt the theoretical relevance as well. When money is treated as endogenous money then the stock of money is determined by the demand for money, and the stock of money itself does not constitute net worth. The theoretical reasons for the stock of money influencing the level of aggregate demand depend on the 'real balance effect' which disappears when money has no net worth.

Another adjustment mechanism, which has recently been more widely used in analysis and to some degree in policy, and which has rather more plausibility than the real balance effect, arises from the operation of interest rate policy by the Central Bank. It should be noted that this is not a market adjustment mechanism, but rather arises from the operation of monetary policy. This could occur if the Central Bank adopted some form of 'Taylor's rule' under which

the setting of the key interest rate depends on the 'equilibrium' rate of interest, deviation of inflation from target and deviation of output from trend level (Taylor 1993). The 'equilibrium' rate of interest is then seen to be that which brings aggregate demand in line with available supply (and a constant rate of inflation).

However, there has to be considerable doubts as to whether feasible variations in the rate of interest are sufficient to equate savings and investment at a high level of economic activity. There are constraints on the extent to which interest rates can be varied (whether for reasons akin to a liquidity trap in operation which prevent the reduction of interest rates below a particular level or for foreign exchange considerations) and there are doubts relating to the potency of interest rates to influence aggregate demand (Arestis & Sawyer 2004b). These considerations are significant in two respects. First, they suggest that there are no automatic market mechanisms that will create sufficient aggregate demand, and hence there is a need for macroeconomic policy to do so. Second, the rate of interest is a possible policy instrument but doubt can be cast on its effectiveness in securing high levels of demand (Arestis and Sawyer 2004a).

The Kaleckian-Keynesian perspective (which may be summarised by the term 'functional finance'; Lerner 1943) is firmly based on two propositions. First, a market economy involves elements of instability, and specifically there are no strong forces leading the economy to full employment. There is no reason to accept Say's Law that 'supply creates its own demand', and in general the level of private aggregate demand will not be compatible with full employment, and usually (but not universal)

private aggregate demand will fall short of full employment. By reference to equation (2a) it can be readily seen that in those circumstances a budget deficit is required to achieve high levels of employment.

Second, the budget position should be used in pursuit of economic policy objectives (notably the achievement of high levels of employment) and not to meet some arbitrary rule (such as balancing the budget over a year or over a business cycle). The 'functional finance' approach is that the purpose of fiscal policy is seen as being to secure the desired level of economic activity, rather than to achieve a particular budget position (e.g. to balance the budget whether on an annual basis or over the course of the business cycle). A budget deficit is used to boost aggregate demand when it would otherwise be insufficient to reach that desired level of economic activity (and also a surplus run when aggregate demand would otherwise be too high).

The general proposition of the 'functional finance' approach is that the budget position should be used to secure a high level of economic activity in conditions where otherwise there would be a lower level of economic activity. Lerner (1943) put the case for functional finance, which "rejects completely the traditional doctrines of 'sound finance' and the principle of trying to balance the budget over a solar year or any other arbitrary period" (p. 355). Functional finance supports the important proposition that total spending should be adjusted to eliminate both unemployment and inflation.

In a similar vein, Kalecki (1944a) argued that sustained full employment "must be based either on a long-run budget deficit policy or on the redistribution of income" (p.135). Kalecki based his argument on the assumption that there would be a tendency

for the level of aggregate demand to fall short of what was required for full employment. Then there was a need for either a budget deficit to mop up the difference between full employment savings and investment, or for full employment savings to be reduced through a redistribution of income from rich to poor. He also argued that ‘although it has been repeatedly stated in recent discussion that the budget deficit always finances itself—that is to say, its rise always causes such an increase in incomes and changes in their distribution that there accrue just enough savings to finance it—the matter is still frequently misunderstood’ (Kalecki 1944b:358). In other words, government spending will generate an increase in income leading to a sufficient rise in savings to fund any excess of government expenditure over tax revenue.

In this approach, a budget deficit is used to boost aggregate demand when it would otherwise be insufficient to reach that desired level of economic activity. From that perspective, the role of fiscal policy should be evaluated against a background of insufficient aggregate demand since the case for fiscal policy is the case to address insufficient aggregate demand. Basing an analysis of the effects of budget deficit on economic activity under the assumption that aggregate demand is always sufficient (i.e. assuming something akin to Say’s Law) is then quite inappropriate: if aggregate demand was sufficient there would be no requirement from a ‘functional finance’ perspective for a budget deficit.

### **Sustainability of Budget Deficits**

The argument is put that long-term budget deficits are unsustainable in that each year’s budget deficit adds to the public debt and

future interest payments. The continuation of a primary budget deficit (that is deficit excluding interest payments) involves the build up of debt and interest payments, and further borrowing to cover those interest payments and the continuing primary deficit. Although in such circumstances, the budget deficit would be growing (when interest payments are included) and the public debt would also be increasing, their relationship with GDP depends on the growth of the economy as well as the level of interest rates. Domar (1944) provided an early analysis of this and saw ‘the problem of the debt burden [as] essentially a problem of achieving a growing national income’ (p. 822), though when his analysis used numerical values for key variables rates of interest of 2 per cent and 3 per cent were assumed. Kalecki (1944b) argued that an increasing national debt did not constitute a burden on society as a whole since it is largely an internal transfer, and further noted that in an expanding economy the debt to income ratio need not rise if the rate of growth is sufficiently high (as further discussed below).

It can be readily shown that a continuing primary budget deficit equivalent to a proportion,  $d$ , of GDP will lead to a debt to GDP ratio stabilising at  $b = d/(g - r)$  (where  $g$  is the growth rate and  $r$  interest rate, either both in real terms or both in nominal terms). Let the outstanding public sector debt be  $D$ , and then the budget deficit is  $dD/dt$  and is equal to  $G + rD - T$  where  $r$  is the post-tax rate of interest on public debt,  $G$  is government expenditure (other than interest payments) and  $T$  is taxation (other than that based on receipt of interest from government). With  $Y$  as national income, we have :

$$\begin{aligned}
& (5) \, d(D/Y)/dt \\
& = (1/Y) \, dD/dt - (D/Y)(1/Y) \cdot dY/dt \\
& = (G + rDT)/Y - (D/Y) \, g
\end{aligned}$$

where  $g$  is the growth of national income. The debt to income ratio rises (falls) if

$$(2c) \, (G-T)/Y > (D/Y)(g-r).$$

It is evident that the stabilisation of the debt to income ratio (with a given primary deficit) requires that  $g > r$ .

In a similar vein, a continuing budget deficit relative to GDP of  $d'$  (including interest payments) leads to a debt to GDP ratio stabilising at  $d'/g$  where here  $g$  is in nominal terms. But this implies that  $b + rd = gd$ , i.e.  $b = (g - r)d$  and hence if  $g$  is less than  $r$  the primary budget deficit is negative (i.e. primary budget is in surplus).

However, in the functional finance approach, the budget deficit which is relevant is the overall budget position rather than the primary deficit (or surplus). To the extent that a budget deficit is required to offset an excess of private savings over investment, then it is the overall budget deficit which is relevant (see below for some caveats). Bond interest payments are a transfer payment and add to the income of the recipient, and similar in many respects to other transfer payments. In terms of sustainability, then, of a fiscal deficit, the condition under 'functional finance' is readily satisfied (with the requirement of growth being positive).

If a budget deficit (particular size relative to GDP) is run for a number of years, then it is clear that the interest payments component of the deficit will increase, and the appearance is given that interest payments are 'crowding' out other forms of public expenditure and/or leading to higher levels

of taxation. However, in the case in which it is the overall budget deficit which is relevant for the level of demand, then a constant total deficit (relative to GDP) can be shown to lead to the convergence of the debt to GDP ratio to the equivalent of the ratio budget deficit divided by the nominal growth rate.

### Globalisation and Budget Deficits

The operation of fiscal policy, often associated with budget deficits, has often been seen subject to the vagaries of the financial markets. Since budget deficits involve borrowing at a price (interest rate on bonds), the financial markets may refuse to lend or to impose a high rate of interest. However, the argument above was that when budget deficits are run to offset an excess of savings over investment, funding for budget deficit would be readily available. From the perspective of the financial markets, the savings can only occur if a budget deficit is operated.

In a world of open capital markets with large financial flows across the exchanges, the argument is made that budget deficits will not be tolerated by the financial markets, and they respond to budget deficits by an outward flow of capital, depreciation of the exchange rate and higher interest rates.

It can first be remarked that the experience of the past decade has been that governments (e.g. USA, Germany) have run substantial budget deficits (of the order of 3 to 4 per cent of GDP) without any apparent difficulties in terms of financing the deficit or effect on the exchange rate. In these examples the budget deficits of the early 2000s ran counter to previous policies (the budget surplus under the Clinton administration in the USA, and adherence to the SGP in the case of Germany). It can be

further argued that financial markets could potentially respond to any set of policies which were not to their liking through withdrawal of money. Arguments such as loss of confidence, poor ‘fundamentals’ would no doubt be much heard.

It is also clear that crises occur for many reasons, generally unrelated to the fiscal policy of the country concerned. Currency and financial crises have been a long-standing feature of capitalism, and a feature which has increased in the past 30 years or so. The East Asian financial crisis of 1997 is an example of a crisis which was unrelated to budget deficits or what may be seen as reckless macroeconomic policy.

## Conclusion

The notion of ‘fiscal consolidation’ (or other expressions such as ‘fiscal stability’) became a hallmark of the Washington consensus on economic policy, and involved measures towards limiting budget deficits and stabilising public debt (relative to GDP). The EU’s Stability and Growth Pact was an example of the influence of this widely held notion on policy. The failure of many EU member states to adhere to the 3 per cent of GDP set for budget deficits illustrates the difficulties and dangers of such a policy. The major task is to re-establish the idea that, in the face of inadequate aggregate demand, budget deficits have a clear role to play in raising the level of economic activity.

## Selected References

Alesina, A., R. Perotti and J. Tavares. (1998) “The Political Economy of Fiscal Adjustments”, *Brookings Papers on Economic Activity*, Number 1, pp. 197-248  
 Arestis, Philip and Sawyer, Malcolm. (2004a) “On the Effectiveness of Monetary and Fiscal Policy”, *Review of Social Economy*, Volume 62, Number 4, pp. 441-463

Arestis, Philip and Sawyer, Malcolm. (2004b) “Can Monetary Policy Affect the Real Economy?”, *European Review of Economics and Finance*, Volume 3, Number 3, pp. 9-32  
 Buiter, Willem H. (2001) “Notes on “A Code for Fiscal Stability””, *Oxford Economic Papers*, Volume 53, Number 1, pp. 1-19.  
 Buti, Marco; Daniele Franco and H. Ongena. (1997) *Budgetary Policies during Recessions: Retrospective Application of the “Stability and Growth Pact” to the Post-War Period*. Brussels: European Commission.  
 Domar, E.D. (1944) “The “Burden of the Debt” and the National Income”, *American Economic Review*, Volume 34, Number 4, pp. 798-827.  
 H.M. Treasury. (1999) *The New Monetary Policy Framework*. London: HMSO  
 H.M. Treasury. (2005) *Budget 2005: Economic and Fiscal Strategy Report and Financial Statement and Budget Report*. London: HMSO  
 Howson, Susan. (1975) *Domestic Monetary Management in Britain 1919-38*. Cambridge: Cambridge University Press  
 Kalecki, Michal. (1944a) “The White Paper on Employment Policy”, *Bulletin of the Oxford University Institute of Statistics*, Numbers 6-8, pp. 131-135.  
 Kalecki, Michal. (1944b) “Three Ways to Full Employment”, in *The Economics of Full Employment: Six Studies in Applied Economics*, Prepared at the Oxford University Institute of Statistics; Oxford: Blackwell. Pages in text refer to Jerzy Osiatyński (Editor), *Collected Works of Michal Kalecki*, Volume 1, *Capitalism: Business Cycles and Full Employment*. Oxford: Clarendon Press.  
 Keynes, J. Maynard. (1980) *Activities 1940-1946 Shaping the Post-War World: Employment and Commodities*, *Collected*



- Writings*. Volume 27 of Collected Works of J.M. Keynes. London: Macmillan.
- Lerner, Abba. (1943) "Functional Finance and the Federal Debt", *Social Research*, 10(1) 38-51; reprinted in W. Mueller (ed.) *Readings in Macroeconomics*. New York: Holt, Rinehart and Winston, pp. 353-360.
- Middleton, Roger. (1985) *Towards the Managed Economy*. London: Methuen
- OECD (2000) *Economic Outlook* 68. Paris: OECD.
- Seidman, Laurence S. (2003) *Automatic Fiscal Policies to Combat Recessions*. Armonk, New York: USA: M.E. Sharpe
- Wren-Lewis, Simon. (2000) The Limits to Discretionary Fiscal Policy, *Oxford Review of Economic Policy*, Volume 16, Number 4, pp.92-105

*Malcolm Sawyer*  
*Leeds University Business School*  
*University of Leeds, UK*  
*mcs@lubs.leeds.ac.uk*

## Business Cycles

*Jan Reijnders*

### **Introduction**

The observation that economic development does not occur smoothly but tends to proceed in jerks and leaps has long fascinated economists. Originally they concentrated on the crisis phenomenon and were inclined to attribute disturbances in the path of economic growth to exceptional circumstances which they assumed to have aroused speculation and reckless behaviour. In the course of the nineteenth century, however, economists became dissatisfied with the interpretation of "*each crisis appearing to be the result of its own separate accident*" (Mills 1867:11). They came to designate recurrent periods of prosperity and depression as wavelike movements (Phillips 1828) or cycles (Lloyd 1837; Wade 1833). They were impressed by the regularity and apparent periodicity of these movements and tended to interpret them as the 'heart-beat' of a living organism. It was expected that the study of this 'heart-beat' would reveal the basic characteristics of the capitalist economy and that it would lead to an understanding of its fundamental laws of motion.

This change in interpretation marks the beginning of the theory of the business cycle. Marx for instance stressed the importance of the study of the "industrial cycle" which he associated with the periodical reproduction of fixed capital (Marx 1893:185-186). In his view, the cycle was one of the manifestations of the historical limitations of the capitalist mode of production. By its periodical recurrence the crisis would, every time more threateningly, put the entire bourgeois

society on trial. Some may dislike the pessimism and determinism of the Marxist view and be suspicious of theoretical constructs related to it. It is however by no means necessary to make doomsday inferences about industrial cycles and diagnose them as pathological.

While sticking to the point of view that the study of cyclical movements is necessary because it reveals the characteristic working of the economic mechanism, one can just as well draw more optimistic conclusions. An instance of this is the impulse and propagation theory of the cycle (Aftalion 1909,1913; Frisch 1933; Pigou 1927) and its present day followers (Friedman & Schwartz 1963; Lucas 1975; Kydland & Prescott 1982; Long & Plosser 1983; Blanchard & Fischer 1989; Blanchard & Quah 1989; Romer 2006). Here cyclical movements are interpreted as manifestations of a mechanism which counterbalances and absorbs the external shocks to which the economy is subjected. They reflect the flexible way in which the economy interacts with its environment. For this school, cyclical movements are not a manifestation of vulnerability or of fatal instability but of vitality and strength.

In between these opposing views is a third interpretation which grants that cycles are inextricably bound to the development of a modern economy and that they present a problem in that they lead to social waste, that is sub-optimal use of scarce resources. The cycles are however not considered an incurable disease. On the basis of a thorough analysis of the mechanism of the cycle, it is possible to develop an effective antidote in the shape of anti cyclical fiscal or monetary policies (Keynes 1936; Myrdal 1939; Tinbergen 1936; Tobin 1980; Basu & Taylor 1999; Tobin 1996).

Despite all the differences between the various views about the fluctuating patterns of economic development they have one important characteristic in common, namely that they acknowledge the importance of the study of these patterns because of the clues they may provide to the working of the economic mechanism as a whole. This is considered to be the relevance of the study of business cycles and it is this point of view which is the starting point and main concern of this paper.

### **From Accidents to Endogenous Factors**

The Classical economists like Adam Smith and David Ricardo conceived the economy as a system that was governed by equilibrating forces that guaranteed that optimal use is made of all available resources. The power of this ‘invisible hand’ is so strong that equilibrium is the rule rather than the exception. On the basis of this one would expect a rather smooth development of the economy. In actual fact the development of the economy of the time was not smooth at all. Instead it was characterised by substantial fluctuations which gave the impression that equilibrium was manifestly absent. The mainstream economists of the time immunised their position by pronouncing that the apparent irregularities were the effects of external disturbances that hit upon the system that would subsequently quickly revert to equilibrium. Each hick-up was explained by reference to some external factor that was made responsible for it and gave it its name: ‘Tulip Mania’, ‘*Kipper- und Wipperzeit*’, ‘South Sea bubble’, ‘Melancholy Decay of Credit’, ‘Mississippi bubble’, ‘Manchester Panic’ etcetera (see Kindleberger 2000; Schumpeter 1939; Wood 1999). It is important to realise that this impression of

things has a direct bearing on economic policy. On this view the economy may occasionally be disturbed by some external ‘accident’ but it contains a powerful internal mechanism that makes sure that it quickly adjusts and promptly returns to equilibrium. For this reason the economy is considered robust and resistant to external shocks by itself. The problem is however that the mechanism that is responsible for this robustness is itself very sensitive of actions that interfere with its workings. State interventions to stabilize the effects of external impulses generally do interfere with the operation of the equilibrating forces and hence turn out to be destabilising rather than stabilising. For this reason there is no room for economic policy because this is assumed to throw grit into the wheels of the equilibrating mechanism.

Accidents can happen and there is always the possibility that some external event will push the economy off its track. The problem is that once such event has run its course, the economy does not immediately revert to its original position. It rather tends to deviate from it in a cumulative fashion. It is precisely this cumulative process and the length of the period in which the economy deviated from its postulated equilibrium position that worried the dissenting economists of the time and for which they tried to give an explanation. Sismondi, for instance, who coined the expression ‘commercial crisis’ in his book of 1819 (Simonde de Sismondi 1819), defined a situation of general glut where the consumption potential of the workers falls short of the supply of goods. The overproduction implies curtailment of production which reduces employment and hence brings down the consumption potential of workers which further

intensifies the overproduction. Similar points of view regarding the importance of (components of) aggregate demand and the possibility of a general overproduction may be found in the work of Lauderdale, who worried about the possible effects of a reduction of government spending after the Napoleonic Wars (Lauderdale 1804) and Malthus, who took issue with Ricardo on the tenability of Say's Law (Malthus 1820). The principle of the workings of such a cumulative process, the leapfrogging of employment and demand, has survived the ravages of time. In a more sophisticated form it still is a basic ingredient in many modern business cycle theories. But as such it is only part of the story. It is capable of explaining the conditions of crisis, or for that matter, which of the characteristics of the capitalist economy make it prone to economic crisis, and of explaining why it persists, that is why a crisis, once set in, may lead to a cumulative downturn (with the ultimate possibility of a complete breakdown) of the system. The problem is that it does explain a cumulative movement in a downward direction but it cannot explain why the economy time and again recovers from this downfall. Neither can it explain the apparent regularity or periodicity of this movement.

When in the course of the 19<sup>th</sup> century more statistical material became available, scholars got impressed by the regularity and apparent periodicity of the movements of the economy and became accustomed to calling them cycles. See figure 1. In 1862 the French economist Clément Juglar presented an extensive analysis of the available statistical material and suggested that commercial crises would recur periodically. He stressed the cumulative effect of the interaction between economic quantities and

defined commercial crisis as one stage in a three phase cycle of prosperity, crisis and depression (*prospérité, crise, liquidation*). He insisted that once started the sequence was driven by an endogenous mechanism wherein each subsequent phase emanates from its predecessor. Juglar's basic weakness was that he did not provide an explanation of the way in which this endogenous mechanism was ignited. Obviously he supposed that some external factor caused a phase of over-optimism which gave rise to an increase in the price level that in turn gave rise to a surge of speculation. So he envisioned an endogenously driven cycle that was periodically renewed by some external impulse in the shape of a sudden price rise.

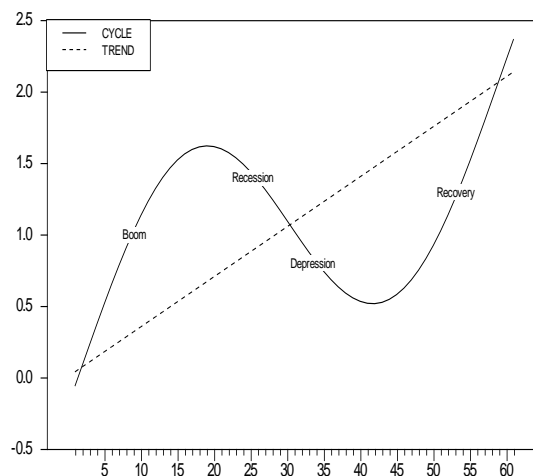


Figure 1 Illustration of Trend plus Four Phase Cycle

On the basis of a substantial empirical analysis, William S. Jevons (1875,1878) advanced a theory of a cyclical movement of the economy based on the periodical occurrence of good harvests which in Jevons's view were precipitated by an 11-years cycle in solar activity (sun spots). Here again we have a logical sequence driven by an endogenous mechanism of which the basic rhythm is determined by some external

factor that acts as the 'metronome' for the movements of the economy at large.

About the same time, Marx introduced an interpretation wherein an endogenous cycle mechanism generates its own impulse for renewal. In his view the industrial cycle is the manifestation of the 'heart beat' of the capitalist economy. The ultimate source of its rhythm must be found in the technical conditions of its reproduction process. The typical feature of modern capitalist production is the extensive use of machinery. According to Marx the stock of fixed capital has a definite life span. A concentration of investment in fixed capital at a certain point in time would therefore produce the concentrated need for replacement investment after the life span of the machinery has passed. Subsequently the original investment impulse is echoed through time to produce a series of impulses for the cyclical mechanism.

### **Endogenous Cycles**

It was the Russian economist Tugan Baranowsky who adopted Marx's analysis of the reproduction process of capital and included the reproduction schemes in his theory of economic crisis (Reijnders, 1997). Social production is subdivided into the contributions of different sectors (means of production, consumer goods and luxury goods) which are interdependent in the sense that means of production are a necessary input for all sectors, that consumption goods are necessary for the reproduction of the labour force of all sectors and, together with the luxury goods, for the reproduction of the capitalists themselves. To maintain an undisturbed process of expansion it is required that these different sectors of production develop in step. If not there will be an overproduction in one of the sectors

that will lead to a curtailment of production in this sector which through its effect on employment and consumption will spread to the other sectors and lead to a general overproduction. The critical task of any system of social production will be to maintain this proportionality in order to avoid economic crisis.

According to Tugan Baranowsky it is precisely at this point where the capitalist mode of production is bound to have problems. In his view the capitalist mode of production is characterised by two fundamental contradictions. The first is the antagonism between production as a means of satisfying human needs and production as a factor in the creation and accumulation of capital. The second is the antagonism between the organisation of production in the individual firm and the anarchy of production in society as a whole.

In view of these contradictions, capitalism will have a difficulty in maintaining the proportionality between sectors. A disproportional division implies a partial overproduction of some commodities, which easily leads to a general glut. Accordingly the two fundamental contradictions of the capitalist mode of production make sure that economic crises become part and parcel of capitalist development.

So Tugan Baranowsky makes the institutional arrangement of capitalism responsible for the occurrence of economic crisis but the next question is how does he explain the recurrence of this phenomenon? In his view the ups and downs of capitalist economy are driven by an endogenous mechanism that explains why an expansionary movement once set in motion will overshoot its equilibrium and produce a crisis. The crisis is the onset of a downturn during which proportionality is restored and

in which the preconditions for a new upturn are fulfilled. The mechanism consists of a mixture of monetary and real factors. An upswing is initiated by the availability of free loanable funds which desperately seek for an outlet in productive investment. Once such an outlet is found there will be an increase in the demand for and the production of means of production which creates extra employment and income and consequently an increase in the demand for consumer goods. The increased demand for consumer goods creates an additional derived demand for means of production. Because the production of means of production takes time (gestation period), the growth of the production of consumption goods and hence the derived demand for means of production outruns actual production of means of production. It is this leapfrogging between the means of production sector and consumer goods sector which creates the path of the upswing.

This process, however, cannot go on for ever. During expansion the reservoir of free loanable funds is gradually used up and the financing of investment encounters an upper limit. The boom is so to speak asphyxiated by the shortage of financial funds. The boom peters out but because the production of means of production cannot be stopped immediately (because of its gestation period) a disproportionality, an overproduction of means of production, occurs. This partial overproduction develops into a general overproduction and a downswing develops. The leapfrogging mechanism is put into reverse and the activity level drops. During the downswing there are two factors at work that prepare the stage for the next upswing. Firstly, because the rate of investment decreases faster than the rate of accumulation of funds, the reservoir of

loanable funds is replenished. Secondly, because the reduction in the activity level of the means of production sector tends to be stronger than the reduction in consumer goods production, the proportionality between sectors is restored. With the correct proportions between sectors restored and the pressure from investment seeking funds building up, the upturn is only a matter of time.

Tugan Baranowsky thus demonstrates that economic crises are recurrent by creating an endogenous theory of the industrial cycle. He compares the fluctuating economy to a steam engine. The loanable funds play the role of steam. The pressure of the steam sets the piston in motion and pushes it to the end of the cylinder. Here the steam escapes and the piston returns to its former position. By analogy the loanable funds set the economy in motion and once they are exhausted the economy returns to its former position where the same sequence starts all over again.

By his construct, Tugan Baranowsky accomplishes two things. First he demonstrates that the economic crisis is not a twist of fate that results from factors outside of the economy but that it is intimately connected to the institutional structure of the capitalist mode of production. Secondly he demonstrates that this crisis is nothing but a fleeting moment in a definite succession of phases, the persistence of which gives it the appearance of a cycle. Once set in motion, it will repeat itself in a similar fashion through time.

It is interesting that on both points Tugan Baranowsky found himself in the middle of a raging battle over policy that became known as the 'breakdown controversy' (Sweezy 1942:190-214). Partly this was due to the fact that Tugan Baranowsky—to justify his disproportionality theory of

crisis—took issue with the Marxists by criticizing their alleged breakdown theories, that is, Marx's law of the falling tendency of the rate of profit and his underconsumption theory of crises (Tugan-Baranowsky 1901:235-246). Partly this was due to the fact that Tugan Baranowsky's propositions perfectly fitted in with the ideas of revisionist currents in the socialist movement of the early 20<sup>th</sup> century. If every economic crises and the corresponding downturn is automatically succeeded by a new upturn, as Tugan Baranowsky asserts, the capitalist system will never break down. So there is time for structural and institutional reform. Moreover, if the ultimate cause of economic crisis lies in the institutional characteristic of capitalism that it lacks coordination at the macro level, the intensity of crises will probably diminish when the degree of anarchy of production is reduced. The revisionist expectation was that coordination at the social level would automatically increase with the reduction of cut-throat competition when industry becomes more and more organized in trusts and with the advance of state supervision over the economy. This comes close to the Social Democratic ideal of the 'mixed' or 'managed economy' that became popular after the Second World War.

### **The Keynesian Connection**

Tugan Baranowsky's (1901) book was very influential for European business cycle theory. It was a principal source of inspiration for Wicksel, Spiethof, Schumpeter, Cassel and Aftalion and it can be linked to the work of Clark in the USA and in the UK to Pigou and to Keynes. The latter commented: *"I find myself in strong sympathy with the school of writers [...] of which Tugan Baranowsky was the first and*

*most original"*. This does not imply that it is all in Tugan Baranowsky but he certainly started a new way of thinking about economic dynamics that has kept its relevance to the present day. There are two principal elements contained in his analysis that prove to be strategic in many variants of business cycle theory:

The first has to do with the cumulative process wherein demand, output, employment, income, and consumption interact to produce derived demand, derived output etcetera. It was described by Tugan Baranowsky in terms of the interdependence of sectoral demands. It was later more precisely formalised by Kahn (1931) and Keynes (1936) and became known as the 'macroeconomic multiplier'.

The second has to do with the interconnection between the demand for means of production (fixed capital goods) and changes in the level of output. This also is a cumulative process but it qualitatively differs from the above mentioned multiplier process because of the effects of the longevity of fixed capital items and of the existence of a gestation period for producing them. Longevity explains the empirical fact that fluctuations in the producer goods industry are stronger than the corresponding fluctuations in the consumer goods industries. The existence of a gestation period explains how decisions of the past play a direct role in present day productive activity. It was used by Tugan Baranowsky (1901) and Spiethoff (1925) to explain the incidence and subsequent disappearance of disproportionality between sectors. Through the work of Aftalion (1913), Clark (1917), Pigou (1927) and Cassel (1932), it became known as the 'macroeconomic accelerator'.

The two cumulative processes and the implied time lags taken together form a

mixture that appears to have explosive tendencies. After all, If a change in investment leads to a cumulative change in income (multiplication) and if a change in income in turn leads to an induced change in investment (acceleration) both mechanism seem to amplify each other thus creating a strong snowball effect. Paul Samuelson however demonstrated that the dynamics of this mixture critically depend on the magnitude of the parameters of the processes involved. In his famous article published in 1939 he presents a small macroeconomic model containing interaction between the multiplier and accelerator processes (Samuelson 1939). It consists of three equations: A consumption function, an investment function and an equilibrium condition. In its reduced form, the model consists of a difference equation of the second order of which the dynamic properties can be precisely determined on the basis of the relative magnitudes of the parameters involved. The type of behaviour can be read from Figure 2, below, which has the magnitude of the propensity to consume ( $\alpha$ ) on the vertical and the accelerator coefficient ( $\beta$ ) on the horizontal axis.

At the one extreme with relatively low values of the propensity to consume ( $\alpha$ ) and especially of the accelerator coefficient ( $\beta$ ) as in region A of Figure 2 (which seems like heaven for the Classical economists), the economy will smoothly convergence to its equilibrium position after an initial disturbance. At the other extreme with relatively high values of the parameters, especially the accelerator coefficient, as in region D of Figure 2 (which seems like hell for the Classical economists) an initial disturbance will cause an explosive movement away from equilibrium in an upward or downward direction depending on

the sign of the initial disturbance.

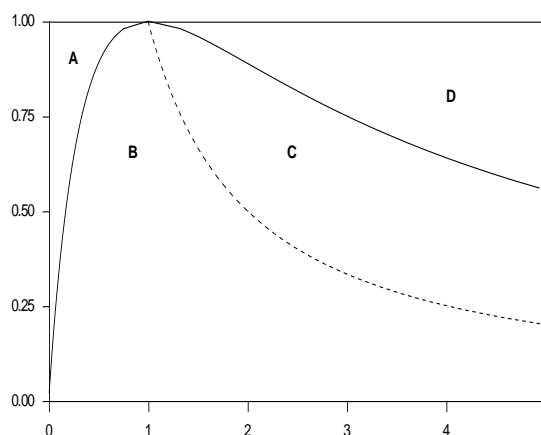


Figure 2 Boundaries of Regions with Different Qualitative Behaviour. Vertical axis:  $\alpha$  = marginal propensity to consume. Horizontal axis:  $\beta$  = accelerator coefficient  
Source: Adapted from Samuelson (1939:78).

For the intermediate values of parameters (regions B and C under the bell-shaped curve in Figure 2) the endogenous cycle comes into its own by producing a cyclical process that may come in the three variants represented in Figure 3, below:

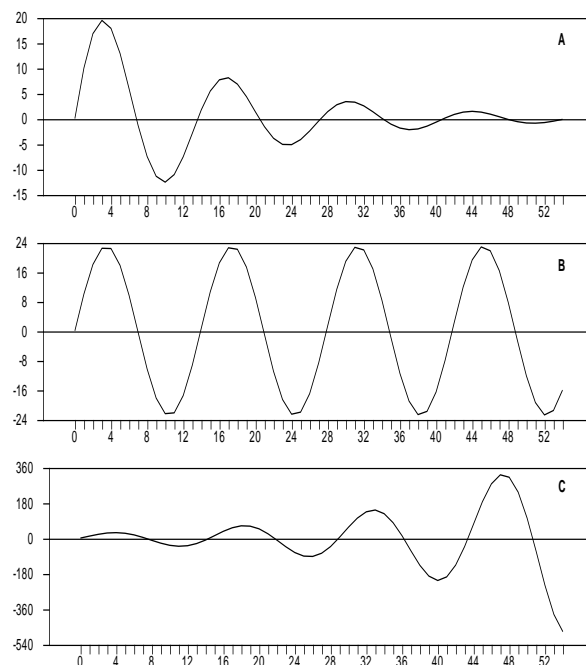


Figure 3 Various Cycle Types:

- A: Damped cycle (shock 10 at  $t=1$ ,  $\alpha=0.8$ ,  $\beta=1,1$ )
- B: Sinusoid cycle (shock 10 at  $t=1$ ,  $\alpha=0.8$ ,  $\beta=1,25$ )
- C: Explosive cycle (shock 10 at  $t=1$ ,  $\alpha=0.8$ ,  $\beta=1,4$ )



Combinations of parameter values in region B produce a damped cycle wherein the amplitude of the cyclical deviation from the equilibrium dies down in the course of time (Figure 3a). Combinations of parameters as in region C will produce the opposite: an explosive cycle wherein the deviation from equilibrium increases with time (Figure 3c). On the (dotted) borderline between B and C the damping and explosive tendencies exactly cancel out thus producing a sinus function with constant amplitude (Figure 3b).

Provided that a characteristic of an endogenous cycle is that, once started, it keeps repeating itself endlessly, only the constellation of parameters on the (dotted) borderline between regions B and C (where  $\alpha=1/\beta$ , that is where the propensity to consume exactly equals to the inverse of the accelerator coefficient) produces a cycle of the requested type. The damped cycle of region B will evaporate in the course of time, whereas the economic system will break down under the influence of the ever increasing amplitude of the explosive cycle of region C. This doesn't bode well for the endogenous cycle hypothesis since it then has to rely on very special assumptions regarding a constellation of parameters which will hardly ever be realised in actual practice. There seem to be two ways out of this problem. Either one adds a constraining factor that keeps the explosive tendencies of the C-type cycle type in check, or one adds an impulse factor that rekindles the B-type cycle once in a while to prevent its amplitude from dying down.

John Hicks, who considered the damped cycle as being inconsistent with historical experience, chose the first option. He introduced a variation of the multiplier-accelerator model based on the proposition

that in reality the value of the accelerator is so high that it produces an explosive cycle. But since historical experience also teaches that the system breakdown that is connected with an explosive cycle did not occur either, there must be a factor that keeps the cycle from breaking outside certain limits. In that case: *"the system might then continue periodically breaking its head against these limits without running away altogether"* (Hicks 1950:92). He assumes that there is an upper limit (a ceiling) to output in the form of limited availability of employable resources (labour supply for instance). See Figure 4, below.

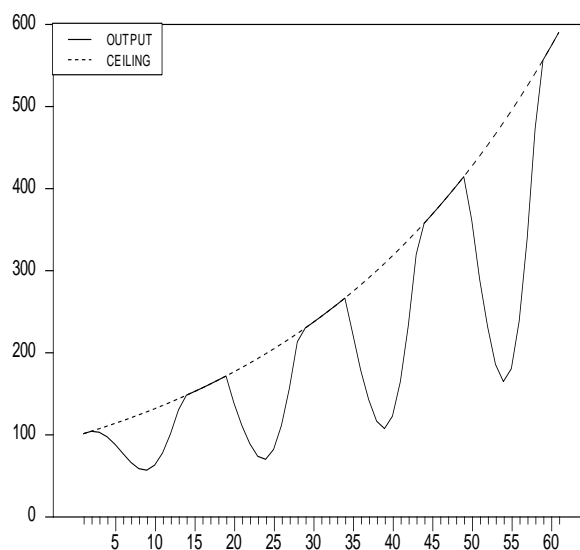


Figure 4 Hicks Constrained Cycle: Explosive Cycle Kept in Check by an Upper Limit to (Human) Resources

Although there is no direct lower limit to

output, Hicks assumption that the accelerator works differently in the downswing provides an indirect check on the decrease in output that will ultimately become effective. Under these conditions an explosive cycle can develop freely as long as its amplitude is lower than the distance between the ceiling and the equilibrium output level. Given the explosive properties of the cycle, however, output will ultimately hit the ceiling. The pace of growth of the upswing can not be maintained. It is forced down to the growth rate of employable resources. Since this rate is lower than the upswing rate, the fall in output growth causes a drop of investment which triggers the downturn that will last until the reduction of productive capacity is no longer sufficient to produce the current level of output. The necessary increase in investment then sets the new upswing in motion.

Samuelson—Hicks type endogenous explanations of the business cycle have become a principal source of inspiration for authors that fit in the Keynesian, Post-Keynesian and (Neo)-Marxist tradition (cf. Duesenberry 1958; Goodwin 1951, 1967; Kaldor 1940, 1951, 1954; Kalecki 1969a, 1969b, 1969c, 1937; Matthews 1959; Laibman 1997; Sherman & Kolk 1996).

### **Econometricians: Exogenous Impulse and Endogenous Propagation**

Hicks view of an explosive cycle that is kept under control by a direct upper limit and an indirect lower limit is one possible solution of the endogenous cycle problem. Another way out is through keeping the damped cycle mechanism intact and adding an impulse mechanism that generates the external shocks which keep the cycle in motion. In this case the damped cycle mechanism is the manifestation of the

response of the economy to the impact of external shocks. It absorbs the displacement by fluctuating back to its equilibrium position analogous to the way in which a pendulum returns to its original position after being pushed away from it.

The standard criticism of the exogenous impulse explanation of economic fluctuations is that the external impulses occur randomly in an irregular fashion. Therefore one would expect that economic time series bear the signs of this by exhibiting a rather erratic pattern. In practice however it appears that the economy tends to fluctuate rather regularly without clear traces of the irregularities reminiscent of the triggering process. One should however make a clear distinction between the impulse and the response mechanism. The first has something to say about the external factors impinging on the system. The second gives the response of the system that is indicative of its structure. Wicksel (1907) put it as follows: *‘If you hit a rocking horse with a stick, the movement of the horse will be very different from that of the stick’*. A more ‘modern’ metaphor for the impulse and propagation mechanism is given by Kendal (Kendall 1948:423). A car is fitted with a suspension system designed to absorb shocks caused by bumps in the road. If the car hits a bump or a pothole in the road, the body of the car will oscillate for a while to absorb the shock, gradually coming to rest. If the car proceeds over a random succession of bumps, it will oscillate continuously. The existence of motion is determined by the sequence of bumps (the impulse mechanism). The regularity of the motion depends on the structure and suspension of the car (the propagations mechanism).

The challenge of explaining how a sequence of random shocks can be

transformed into a rather smooth cyclical fluctuation was taken up by Eugen Slutsky in 1927 (Slutsky 1927, 1937). He put a long series of winning numbers of the Russian State lottery in a row and created a new series consisting of moving sums of 10 consecutive digits (so he added digits 1-10, 2-11, 3-12 etcetera). As such the digits form a series of unrelated random numbers which does not present a particular pattern. If they are transformed into the series of moving sums, however, consecutive numbers are serially correlated and appear to follow a particular pattern. Slutsky then demonstrated that the fluctuations in his series of moving sums bear a striking resemblance to the movements in an index of English business cycles (Slutsky 1937:110). This might indicate that the movements in the English business cycles are of the same origin as the fluctuations in the moving sum series. They might both be the result of applying a linear operator to a series of random numbers.

This idea is taken up by Ragnar Frisch who argues that business cycle theory is confronted with two main problems (Frisch 1933:171). The first is the so-called propagation problem that is the difficulty of explaining the structural properties of a given economic system and the characteristics of the swings they generate once it is set in motion (the shape and the characteristic motions of the rocking horse in Wicksel's metaphor). The second is the impulse problem, the difficulty of identifying the factors that propel the system and keep it moving (the hitting with a stick that keeps the rocking horse in motion). To answer these questions Frisch sets out to create a macro dynamic model capable of explaining the basic movements of the economy. In its simplest form it is expressed as a system of two equations in

two unknowns (consumption and production). This gives a reduced form equation in the shape of a single variable linear differential equation which produces a secular trend but no fluctuations. Next he introduces Aftalion's gestation period of capital to create the possibility for oscillations. The gestation period introduces persistence in the system because it implies that today's activity is affected by decisions in the past.

To study the nature of the solutions to the resulting mixed system of difference and differential equations, Frisch inserts plausible values for its parameters and calculates the time pattern of development starting from an arbitrary set of initial conditions. Now the system generates on top of the earlier mentioned secular trend three distinct cycle types of which the duration mainly depends upon a parameter that is determined by the length of the gestation period of capital. It can be demonstrated that this system is capable of generating cyclical fluctuations. It appears, however, that with the given set of plausible parameters the swings tend to die down in the course of time. According to Frisch this is generally not the case in reality, so he also sees himself confronted with the earlier mentioned endogenous cycle dilemma already addressed by Hicks (Hicks 1950:90,ff). His model only explains the propagation process, but the impulse problem remains. In his view this could be solved by following Knut Wicksel's hypothesis that erratic shocks provide the energy which maintains economic cycles (Frisch 1933:30). To demonstrate how this could be explained he simplifies his model by assuming that it behaves as a swinging pendulum. It is known that its oscillations will gradually die down and the pendulum

will again come to rest in its equilibrium position. Frisch then calculates the path of the pendulum when it is subjected to a sequence of erratic shocks of various strengths exerted in both directions. Next he shows that the ordinate of the pendulum at any moment will be a weighted cumulation of the effects of past shocks. The system of weights is determined by the properties of the pendulum itself. Simulation shows that the shape of the curve is a distorted harmonic with the same frequency as the one that is typical of the pendulum. The result again bears a striking resemblance to the fluctuating patterns that are typical of economic time series. In the example the pendulum is just a metaphor for the economic system itself that is assumed to have similar properties. Accordingly Frisch concludes that: *“by connecting the two ideas: (1) the continuous solution of a determinate dynamic system and (2) the discontinuous shocks intervening and supplying the energy that may maintain the swings”* (Frisch 1933:34) we can provide an explication for the familiar patterns discernible in economic time series.

## Policy

Although the ‘pure’ endogenous cycle theory and the impulse and propagation theory of the cycle provided quite different answers to the ‘endogeneity problem’, their central tenets can, in a technical sense, easily be reconciled. The protagonists of the endogenous cycle theory don’t have a difficulty in accepting that exogenous factors do interfere with the basic oscillating pattern in a random fashion, whereas the impulse and propagation theorists readily accept that there is an internal mechanism that absorbs such external shocks by fluctuating around its equilibrium level, just

like the tree that bends over and springs back to surmount the forces of wind. The consensus is that there is an oscillating mechanism that is characteristic of a modern economy and that can be known. If the mechanism is known it is possible to assess how it will react to a certain disturbance. On the basis of this knowledge one could possibly conduct economic policy if one could identify external impulses and if one would be able to give (counter)impulses oneself. In the era of the Great Depression, which coincided with the heydays of business cycle theory and with the period wherein the economic profession gradually came to the conclusion that the stability of the economy could not be *“safely left in private hands”* (Keynes 1936:320) this idea was further elaborated.

Instead of the simple theoretical models that, fed with plausible values for their parameters produced reasonable mimics of true world business cycles, the objective became to design macroeconomic models that capture the basic properties of the actual economy. By filling it with statistical estimates of the parameters it would then be possible to simulate the behaviour of the economy and to analyze the impact of external impulses. Building on the same type of impulse and propagation ideas as Ragnar Frisch, Jan Tinbergen was the first to build such a model (Tinbergen 1936) thus setting the first step in what became standard practice in many countries around the world. Such a model can be used as a means of making predictions of how the future development of the economy will be. It can also be used to simulate how the economy will respond to external shocks. If one knows how the system responds to shocks one may on the basis of the same model try to stabilize the effects of a major shock by

giving it a (series of) policy shocks as an antidote to the cyclical response to the disturbance. Similarly, the macroeconomic model might be used as a means of uncovering the basic weaknesses of the system that are responsible for excessive instability and to determine whether they can be remedied by the appropriate policy measures.

The period after the Second World War witnessed the establishment of many planning agencies that built large scale macroeconomic models of national economies, or even of the world economy at large, that formed the basis for the prediction of future development and as a guide for the creation of impulse time-tables for economic policy. In this ‘optimal control’ view, the economy was considered a sort of engine with a known mode of operation that could be adjusted by turning the appropriate knobs of its control panel. In this way the economy could be steered in a certain direction and its fluctuations could be mitigated by an explicit anti cyclical policy. The models were mainly built around one or the other version of Hicks’s interpretation of the Keynesian system (i.e. the open economy version of IS/LM Hicks 1937; Fleming 1962; Mundell 1963) and were strongly demand-oriented. Impulses were assumed to be random demand impulses in the exogenous variables that could be countered by an exactly pointed burst of policy measures in order to stabilize their effects. See Figure 5, below.

According to the standard setup of the pendulum model, three types of factors are involved in the actual functioning economy. First is the type of exogenous factors that impinge upon the economic mechanism and that follow an erratic random process. Second is the set of interacting endogenous

factors that represent the internal workings of the economic mechanism and that form the heart of the econometric model. They generate the oscillations of the system and thus assume the role that the pendulum plays in Frisch’s example. Thirdly there is the category of policy variables, a set of monetary and fiscal instruments that is used by the governing institution to stabilize the cycles.

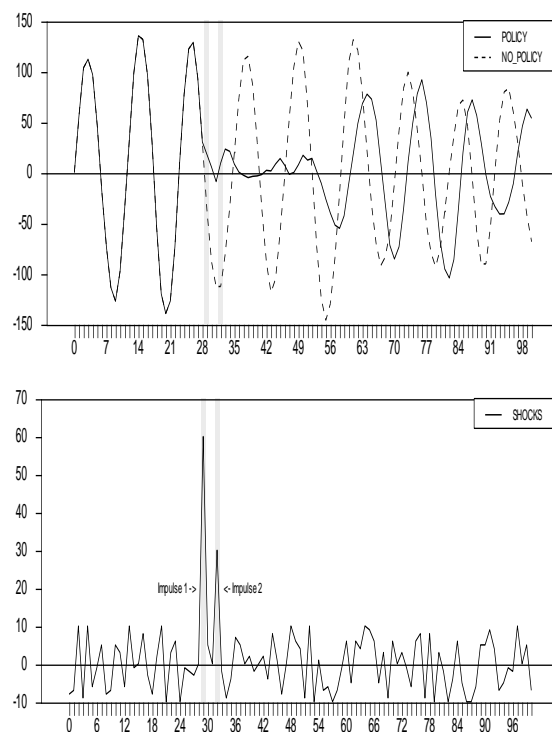


Figure 5 Example of (temporary) stabilization of business cycles by means of policy impulses. NO\_POLICY = cycle without policy. POLICY = cycle after policy impulses at  $t=30$  and  $t=33$ , shocks = random shocks that fuel the cycle)

The big question, however, is whether a given macroeconomic model is capable of adequately reproducing the cyclical behaviour of a modern economy. This is the issue taken up by Irma and Frank Adelman with their analysis of the dynamic properties of the well-known Klein-Goldberger model for the USA (Adelman & Adelman, 1959). On first inspection of their simulation experiment they find that there is no “hint

*whatever of any internally generated business cycle, and, indeed, even in the first few years, the shock of start-up is not sufficient to induce more than a single turning point in any variable"* (Adelman & Adelman 1959:602). If one wishes to maintain that cyclical movements are self-generated one might have to conclude that the Klein-Goldberger model is totally inadequate unless one is prepared to look elsewhere for the origin of the factors fuelling business fluctuations. This is precisely what the Adelmans intend to do. They follow the path set out by Slutsky, Frisch and Tinbergen and explore the sensitivity of the Klein-Goldberger model to the impact of random disturbances. They distinguish two types of shocks. Type I is the sequence of random disturbances exerted on the exogenous variables of the model as in the standard setup of the impulse and propagation model. Type II is a random process in the error term that is added to each individual equation of the model itself.

Simulations reveal that the Klein-Goldberger model driven by a type I random process exhibits cyclical behaviour comparable to the real world cycles of the US economy be it that their amplitudes are much smaller than the actual values. The performance of Klein-Goldberger driven by type II errors is much better. After assessing the results by means of the NBER method of measuring business cycles (see Burns & Mitchell 1946) it becomes obvious that the model produces cycles that are very similar in respect to their average length, the duration of expansions and contractions, the degree of clustering of peaks and troughs, the proportion of specific cycles that are leading or lagging, and their indices of conformity to that of the empirical material for the USA.

All this seems to underpin the conclusions of the pendulum model. But in fact the Adelman conclusions imply two types of change. The first is that the range of possible perturbations is greatly enlarged because now it is not only the exogenous variables but also the endogenous variables that are affected by chance. The second is that, precisely because the endogenous factors are affected by external disturbances, the relations and interaction patterns between the endogenous variables also change. The model no longer mechanically turns out the same results and no longer responds to policy stimuli in the same fashion. Its mode of operation changes because the proportions within the system change under the impact of external influences. The upshot of all of this is that an assessment of the relative importance of the impulse system increases to the detriment of that of the propagation system and that the estimated power of economic policy diminishes because it has to be implemented in a more complex environment.

### **Turning Tides**

Granted that the relative importance of the impulse mechanism increased, the decade following the Adelman paper witnessed a discussion focusing on the question which type of impulses were more prominent, which part of the transmission within the propagation mechanism was the strongest and which type of economic policy was the most effective. The mainstream (fiscalist) Keynesians mainly held real demand factors responsible for economic fluctuations and regarded, because of the magnitude of the corresponding multipliers, fiscal policy as the most effective instrument for stabilizing them. This interpretation was challenged by Milton Friedman who turned the fiscalist

Keynesian view upside down. In his analysis monetary factors rate highest among the impulses, monetary transmission is the dominant propagation mechanism and because the monetary multipliers are the largest, monetary policy is a more powerful tool (Friedman & Meiselman 1963; Friedman & Schwartz 1963). He, however, immediately adds that discretionary monetary policy should not be tried. Due to long and variable time lags it is impossible to ensure that the impact of monetary policy measures arrives at the right moment and it may turn out that monetary policy in the end is pro-cyclical in stead of anti-cyclical (Friedman 1948). He even goes as far as claiming that much of the observed fluctuations must be attributed to misdirected monetary policy actions. For this reason he suggested that monetary authorities should follow a fixed monetary rule that accommodates the process of economic growth in stead of interfering with it.

Notwithstanding the differences of opinion with regard to the effectiveness of economic policy and the different emphasis on the impact of fiscal and monetary policy measures, the Keynesian mainstream and Monetarism had much in common. They used the same short run model (Mundel Fleming extension of IS/LM) the main difference being the emphasis on real versus monetary factors and the fact that the Keynesians tend to explain cyclical persistence with reference to wage and price rigidities, whereas Monetarist tend to explain persistence by money illusion and the slow adjustment of price expectations (adaptive expectations).

The balance of power between Keynesian and Monetarist economists tipped in favour of the last mentioned in the course of the

1970's. One of the central tenets of Keynesian economic policy making, the Phillips curve, came under attack (Friedman 1968; Phelps 1968). After the Oil shock of 1973 - a 'supply shock' that is alien to the Keynesian focus on 'demand shocks' - it appeared that Keynesianism could no longer explain the problems of the time because it lacked an adequate theory of inflation. The replacement of the adaptive expectations rule with the rational expectations hypothesis led to the policy ineffectiveness proposition which maintains that economic policy is actually impossible because rational economic agents anticipate the effect of economic policy and adjust so as to neutralize it (Sargent & Wallace 1975, 1976). Robert Lucas launched his critique of the optimal control theory that lay at the root of large scale econometric model building in the Keynesian tradition. He thought that the assumption that the parameters of the model remain the same when there is a change in policy can not be maintained when expectation formation is rational.

Rational agents will adjust to policy changes by changing their behaviour which in turn changes the mode of operation of the model itself (Lucas 1976). Rational expectations together with the assumptions regarding the intertemporal substitution of labour leaves only room for the impact of external causes when they are not anticipated. In Lucas's equilibrium 'monetary surprise' theory of the business cycle, the economy only deviates from equilibrium when an unexpected external disturbance occurs. Likewise, a monetary shock administered by the monetary authorities to stabilize the economy only has a real effect when it takes the private sector by surprise. Lucas's theory (sometimes referred to as Monetarism mark II, or New-

Classical Theory mark I) is similar to Friedman's in the sense that it is driven by monetary shocks and is equipped with strong equilibrium tendencies. The main difference is that the range of shocks with real effects is limited to unanticipated ones and the persistence of cycles is reduced because rational expectations imply a much shorter adjustment process after a disturbance. Moreover, the assumption of an underlying pendulum process is given up altogether (Chatterjee 2000).

In spite of the sometimes fierce debates between Keynesians, Monetarists and (type I) New-Classicals there was a basic consensus on a number of important points (Snowdon & Vane 2005:330). First, they all considered business cycles as temporary deviations from a smooth underlying trend (this is the trend stationarity hypothesis. cf. Blanchard & Fischer 1989:12)). Second, all considered economic fluctuations socially undesirable because they reduced economic welfare, and third they all considered monetary factors to be important for explaining business cycles. These pillars of consensus were, however, demolished in the course of the 1980s with the rise of a different brand of New-Classical theories which come under the title 'Real Business Cycle Theory (RBC)' (Hartley et al 1998; Kydland & Prescott 1982; Long & Plosser 1983; Prescott 1986). The adjective 'real' is added because the protagonists of this school consider money as super neutral. Monetary impulses lead to immediate price adjustments and hence have no real effects.

The economy only responds to supply shocks (changes in production technology, changes in the environment, energy crises, war and political upheaval, labour unrest, and government regulations which damage incentives) of which productivity shocks are

the most prominent. Because RBC theory wants to integrate the theory of fluctuations into the theory of economic growth, they consider the business cycle an anathema. In their view there is no distinction between trend (the locus of equilibrium) and cycles (the deviation of actual values from equilibrium). Every supply shock dislocates the equilibrium itself and has a permanent effect. There is no such thing as a trend plus cycles; the upward tending and fluctuating pattern that is observed in economic time series is nothing but a random walk with drift (Nelson & Plosser 1982). See Figure 6, below.

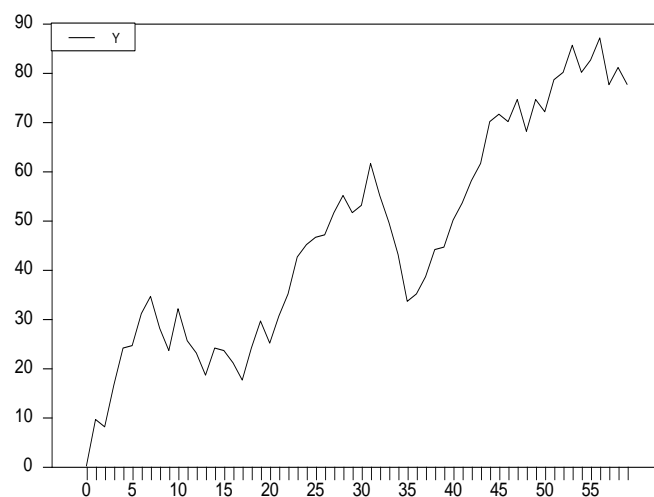


Figure 6. Random Walk with Drift:  
 $y_t = y_{t-1} + \varepsilon_t$  (where  $\varepsilon_t$  = random shock)

The RBC world is populated with representative agents, rationally-expecting intertemporal utility maximizers, who are sensitive to changes in real wages and real interest rates. An increase in labour productivity leads to an upward shift of labour demand. The concomitant increase in real wages is an incentive for the representative agent to increase her labour supply. She substitutes present working hours for future leisure, 'makes hay when the sun shines' if the shock is considered transitory or completely readjusts her assets and activity portfolio if the shock is



considered permanent. According to the RBC theorists, intertemporal substitution is a powerful propagation mechanism. It is, however, not the only one. Consumption smoothing and the gestation period of fixed capital ('time to build', Kydland & Prescott 1982) are also factors that carry forward the impacts of initial impulses.

RBC theorists consider the Solow residual—the part of the growth process that can not be explained by the change of factor inputs—as the principal indicator of technological change. They then demonstrate that feeding this residual to a calibrated version of the RBC-model results in patterns that strongly resemble the ones that are present in some important time series data relating to the US economy (Plosser 1989). So here again, we have a version of the impulse and propagation model of the business cycle that is capable of mimicking the behaviour of an actual economy. But there is an important difference with the older versions of the impulse and propagation approach. The RBC model is an equilibrium model that does not represent the way in which the economy finds its way back to an equilibrium position but rather the way in which an economy moves under the impulse of an external force while constantly remaining in equilibrium.

The upshot of this is a completely different view of the role of economic policy. According to RBC theory the movements of the economy are just the result of rational economic agents responding in an optimal way to changes in the environment. On this view markets always clear and accordingly observed fluctuations can not be interpreted as socially undesirable deviations from an ideal growth path of output. They are considered

to represent Pareto-optimal equilibrium positions. The idea that government would try to reduce them would certainly lead to an undesirable reduction of welfare. Prescott put this as follows: "*the policy implication of this research is that costly efforts at stabilization are likely to be counter-productive. Economic fluctuations are optimal responses to uncertainty in the rate of technological progress*". On this view it appears that the business cycle no longer presents a problem.

### **Resuscitating Business Cycle Theory**

As far as business cycles are concerned it seemed as if mainstream economic theory had definitely reverted to its original 19<sup>th</sup> century position. The omnipresent equilibrium pushed out the business cycle and relegated it to the domain of the type of essentially exogenous processes that are beyond the reach of economic policy. However, as so often in the history of economic thought it was economic reality itself that turned the tables on RBC-theory. The equilibrium explanations of the business cycle were called into question by the recessions of the early 1980s and 1990s and the persistence of unemployment particularly in Europe in the same period (Tobin 1992; Arestis & Sawyer 1998; Snowden & Vane 2005:359). It proved impossible to square the observation of sustained double digit unemployment rates with RBC's notion of full employment equilibrium and the corresponding conception of 'voluntary unemployment' resulting from rational economic agents' choice to substitute work for leisure. They should rather be seen as the expression of involuntary unemployment that is attached to macroeconomic instability resulting from aggregate demand disturbances.

In other words the persistence of high unemployment indicates that there are prolonged periods wherein the economy can be off its full employment equilibrium. Accordingly the facts indicate that there is scope for a return to the earlier mentioned Keynesian-Monetarist consensus that economic development can deviate from its equilibrium trend, that these deviations are socially undesirable and that there is scope for government intervention to improve macroeconomic stability and economic welfare.

The new interpretation, dubbed New Keynesian economics, builds on RBC-theory in the sense that it accepts several of its methodological premises. It agrees that macroeconomic theories require solid micro foundations, that choices are guided by rational expectations and that the general equilibrium framework is best suited for macroeconomic model building. New Keynesians also accept that supply shocks are important determinants of macroeconomic dynamics. They immediately add, however, that demand shocks are as important, particularly so because they define the deviations from (supply-determined) path of equilibrium (Blanchard & Fischer 1989; Blanchard & Quah 1989, 1993; Mankiw 2006).

The essential difference between RBC- and New Keynesian theories relates to their interpretation of the micro foundations of macroeconomics. RBC refers to a notional system of instantaneous adjustment within a full information perfect competition economy with a complete set of markets. They presuppose that all markets clear and hence take the existence of equilibrium for granted. On the other hand New Keynesian theory refers to a real economy that is riddled with imperfections that slow down

the adjustment process so as to create prolonged periods of deviation from equilibrium. The main focus of New Keynesian theorizing is on imperfections in financial markets (Greenwald & Stiglitz 1993), goods markets (Akerlof & Yellen 1985; Mankiw 1985, 2006; Parkin 1986; Greenwald & Stiglitz 1991; Stiglitz 1999) and labour markets (Akerlof & Yellen 1986; Fischer 1977; Katz 1986, 1988; Taylor 1980; A. Weiss 1990; L. Weiss 1986; Yellen 1984) that lead to wage/price rigidities which hamper the operation of equilibrating forces and thus create the possibility for sustained deviations from equilibrium. In this way they reintroduce the typical response time that on the one hand explains why the effect of demand shocks tends to persist and on the other hand creates a window of opportunity for macroeconomic policy.

The New Keynesians again turn the attention to the propagation mechanism inherent in the economic system itself. Persistence is not a property of the impulse mechanism as in RBC-theory (Cogley & Nason 1995) but rather a consequence of the propagation mechanism that carries the effects of shocks through time. In this way the New Keynesians again shift the relative weight from exogenous impulses to the endogenous propagation mechanisms that are at the heart of business cycles. In doing so they also demonstrate that there is scope for macroeconomic policy: The stretch of time in which the economy is off its full employment equilibrium is considered a loss to society (also see Basu & Taylor 1999; Woodford 2003; Zarnowitz 1998, 1999). As far as economic policy can mitigate such losses or shorten the period wherein these are incurred, it must be considered a blessing to society.

This does, however, not imply that the clock is turned back to the euphoria of the 1960s and the early 1970s. The window of opportunity for economic policy is considered to be much smaller than before. The expected number of complications has increased considerably. Firstly, if the economy is assumed to be affected jointly by demand shocks (with a transitory effect) and supply shocks (with a permanent effect), economic policy making is like aiming at an erratically moving target. Stabilizing the effect of a demand shock becomes more difficult when equilibrium is randomly shifted due to supply shocks. Secondly, the assumption of rational expectations implies that the adjustment process is much faster. Accordingly the time frame for policy intervention is much smaller. Thirdly, the assumption of rational expectations implies that policy shocks only have a real effect if they are unforeseen. Under these conditions economic policy making becomes a strategic game in which the monetary and fiscal authorities can only be successful if they take private agents (who do the utmost to try and predict such policy actions) by surprise (Blackburn 1987; Kydland & Prescott 1977). Fourthly, in view of the uncertainties surrounding the movements of the economy and the existence of inside and outside time lags that delay the impact of policy actions it may well be that economic policy is in effect destabilizing rather than stabilizing. 'Fine tuning' the economy may thus be an illusion so that one has to restrict oneself to, what Lindberg referred to as 'coarse tuning', policies designed to counteract or circumvent serious macroeconomic problems (Lindbeck 1992:231).

## Conclusion

In this paper the development of business cycle theory is presented as the vicissitudes of a school of thought that started as a dissenting current opposing the Classical view that economic fluctuations should be considered the result of 'accidents' that only temporarily interrupt the rule of equilibrium. It was the apparent regularity of the occurrence of such accidents that gave way to the interpretation that fluctuations are the manifestation of some fundamental property of the economic system itself. At first this led to the development of a pure endogenous cycle theory but gradually the idea emerged that the observed patterns were brought about by a combination of factors wherein endogenous mechanisms as well as exogenous 'accidents' have a role to play. The insight that the internal mechanism tends to respond to external stimuli brought about the idea that the economy could be guided by autonomous impulses. In this case government could try to mitigate the effects of adverse shocks by stabilizing the movements of the economy.

In the heydays of Keynesianism it was thought that the business cycle could be kept under control. Monetary and fiscal authorities could stabilize it by administering a precisely pointed burst of positive and negative demand impulses. Their actions should be based on an econometric model that adequately described the functioning of the cyclical mechanism. In practice it appeared that large scale econometric models only had a limited capability to generate cyclical patterns. In order to mimic actual cyclical patterns the models had to be 'fuelled' by a constant stream of exogenous impulses. The effect was that interest gradually shifted from the endogenous propagation mechanism to the exogenous impulse mechanism. Even more

so because the real economy proved to be sensitive of supply side impulses (such as the oil-shocks of the nineteen seventies) that had been neglected before.

It is significant that the external impulses, Mill's 'accidents', not only had a comeback as explanatory factors. In the course of time their importance increased relative to that of the endogenous factors to the degree that the former came to dominate the latter. This is the case with RBC-theory where random fluctuations in the rate of technological progress completely determine where the economy is going. Equilibrium is immediately re-established after every disturbance. Observed fluctuations are nothing but a continuously moving equilibrium and nothing can or should be done about it. Accordingly business cycle theory seemed to have gone full circle: The invisible hand seemed to have returned with a vengeance.

The reality of prolonged periods of high unemployment, however, could not be captured in terms of what Tobin called the 'elegant fantasies' of the 'Robinson Crusoe macroeconomics' of RBC-theory (Tobin 1996). Neither could the recessions of the early 1980s, the early 1990 and the recession after 2001 be squared with the idea that that it is only random supply shocks that govern economic life. The pendulum swung back in the direction where the endogenous forces of adjustment and the consequent windows of opportunity for economic policy take centre stage again.

### **Selected References**

Adelman, I. (1960) "Business Cycles - Endogenous or Stochastic?", *Economic Journal*, Volume 70 (December), pp. 783-796.

Adelman, I. (1962) "Business Cycles - Endogenous or Stochastic? - A Reply", *Economic Journal*, Volume 70 (June), pp. 469-470.

Adelman, I. and F.L. Adelman. (1959) "The Dynamic Properties of the Klein-Goldberger Model", *Econometrica*, 27(4), pp. 596-625.

Aftalion, A. (1909) "La realite des surproductions générales. Essai d'une théorie des crises générales et périodiques", *Revue d'economie politique*, 22, pp. 81-117.

Aftalion, A. (1913) *Les crises périodiques de surproduction*. Paris: Marcel Rivière.

Akerlof, G.A. and J.L. Yellen. (1985) "A Near-rational Model of the Business Cycle with Wage and Price Intertia", *Quarterly Journal of Economics*, Volume 100, Number 5, pp. 823-838.

Akerlof, G.A. and J.L. Yellen. (1986) *Efficiency Wage Models of the Labor market*. Cambridge; New York and Melbourne: Cambridge University Press.

Arestis, P. and M. Sawyer. (1998) "Keynesian Economic Policies for the New Millennium", *Economic Journal*, Volume 108, Number 446, pp. 181-195.

Basu, S. and A.M. Taylor. (1999) "Business Cycles in International Historical Perspective", *Journal of Economic Perspectives*, Volume 13, Number 2, pp. 45-68.

Blackburn, K. (1987) "Macroeconomic Policy Evaluation and Optimal Control Theory: A Critical Review of Some Recent Developments", *Journal of Economic Surveys*, Volume 1, Number 2, pp. 113-148.

Blanchard, O.J. and S. Fischer. (1989) *Lectures on Macroeconomics*. Sixth Edition. Cambridge MA: MIT Press.

- Blanchard, O.J. and D. Quah. (1989) "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, Volume 79, Number 4, pp. 655-673.
- Blanchard, O.J. and D. Quah. (1993) "The Dynamic Effects of Aggregate Demand and Supply Disturbances: Reply", *American Economic Review*, Volume 83, Number 3, pp. 653-658.
- Burns, A.F. and W.C. Mitchell. (1946) *Measuring Business Cycles*. New York: National Bureau of Economic Research.
- Cassel, G. (1932) *Theory of Social Economy*. New York: Harcourt, Brace & Co.
- Chatterjee, S. (2000) "From Cycles to Shocks: Progress in Business Cycle Theory", *Business Review*, pp. 1-11.
- Clark, J. M. (1917) "Business Acceleration and the Law of Demand:", *Journal of Political Economy*. Volume 25, Number 3, pp. 217-235.
- Cogley, T. and J.M. Nason. (1995) "Output Dynamics in Real-Business-Cycle Models", *American Economic Review*, Volume 85, Number 3, pp. 492-511.
- Cooley, T.F. (1997) "Calibrated Models", *Oxford Review of Economic Policy*, Volume 13, Number 3, pp. 55-69.
- Coppock, D. J. (1962) "Business Cycles - Endogenous or Stochastic? A Comment", *Economic Journal*, Volume 70, December, pp. 458-468.
- Duesenberry, J. S. (1958) *Business Cycles and Economic Growth*. New York & London: McGraw-Hill.
- Engels, F. (1882) *Die Entwicklung des Sozialismus von der Utopie zur Wissenschaft*. (Vol. 19, Marx & Engels Werke) Berlin: Dietz, 1973.
- Fischer, S. (1977) "Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule", *Journal of Political Economy*, Volume 85, Number 1, pp. 191-205.
- Fleming, J. M. (1962) "Domestic Financial Policies under Fixed and under Floating Exchange Rates", reprinted in M. Ugur (Editor), *An Open Economy Macroeconomics Reader*. London and New York: Routledge, 2002.
- Friedman, M. (1948) "A Monetary and Fiscal Framework for Economic Stability", *American Economic Review*, Volume 38, Number 3, pp. 245-264
- Friedman, M. (1968) "The Role of Monetary Theory", *American Economic Review*, Volume 58, Number 1, pp. 1-17.
- Friedman, M., & Meiselman, D. (1963) "The Relative Stability of Monetary Velocity and the Investment multiplier in the United States, 1897-1958", in *Commission on Money and Credit: Stabilization Policies*. Englewood Cliffs: Prentice Hall.
- Friedman, M. and A.J. Schwartz. (1963) *Monetary History of the United States, 1867-1960*. Princeton: Princeton University Press.
- Frisch, R. (1933) "Propagation Problems and Impulse Problems in Dynamic Economics", in *Economic Essays in Honour of Gustav Cassel*. London: Allen and Unwin, pp. 171-205.
- Goodwin, R.M. (1951) "The Non-Linear Accelerator and the Persistence of Business Cycles", *Econometrica*, Volume 19, Number 1, pp. 1-17.
- Goodwin, R.M. (1967) "A Growth Cycle", in E.K. Hunt & J.G. Schwarz (Editors), *A Critique of Economic Theory*. Harmondsworth: Penguin, pp. 442-449.
- Greenwald, B.C. and J.E. Stiglitz. (1991) "Examining Alternative Macroeconomic Theories", in E. S. Phelps (Editor), *Recent*

- Developments in Macroeconomics. Volume 3. International Library of Critical Writings in Economics, Number 13.* Aldershot, U.K and Brookfield, Vt.: Edward Elgar, pp. 335-388.
- Greenwald, B.C. and J.E. Stiglitz. (1993) "Financial Market Imperfections and Business Cycles", *Quarterly Journal of Economics*, Volume 108, Number 1, pp. 77-114.
- Hartley, J.E., K.D. Hoover and K.D. Salyer. (1998) *Real Business Cycles. A Reader.* London and New York: Routledge.
- Hicks, J.R. (1937) "Mr. Keynes and the "Classics"; A Suggested Interpretation", *Econometrica*, Volume 5, Number 2, pp. 147-159.
- Hicks, J. R. (1950) *A Contribution to the Theory of the Trade Cycle.* Oxford: Clarendon Press.
- Jevons, W.S. (1875) "Influence of the Sun-Spot Period on the Price of Corn", *Nature*.
- Jevons, W.S. (1878) *The Periodicity of Commercial Crises and its Physical Explanation.* London: British Association.
- Kahn, R.F. (1931) "The Relation of Investment to Home Unemployment", *Economic Journal*, Volume 41, Number 162, pp. 173-198.
- Kaldor, N. (1940) "A Model of the Trade Cycle", *Economic Journal*, Volume 50, Number 197, pp. 78-92.
- Kaldor, N. (1951) "Mr. Hicks on the Trade Cycle", *Economic Journal*, Volume 61, Number 244, pp. 833-847.
- Kaldor, N. (1954) "The Relation of Economic Growth and Cyclical Fluctuations", *The Economic Journal*, Volume 64, Number 253, pp. 53-71.
- Kalecki, M. (1937) "A Theory of the Business Cycle", *The Review of Economic Studies*, Volume 4, Number 2, pp. 77-97.
- Kalecki, M. (1969a) "The Business Upswing and the Balance of Payments", in M. Kalecki (Editor), *Studies in the Theory of Business Cycles 1933-1939.* London: Basil Blackwell.
- Kalecki, M. (1969b) "The Mechanism of the Business Upswing", in M. Kalecki (Ed.), *Studies in the Theory of Business Cycles 1933-1939.* London: Basil Blackwell.
- Kalecki, M. (1969c) "Outline of a Theory of the Business Cycle", in M. Kalecki (Editor), *Studies in the Theory of Business Cycles 1933-1939.* London: Basil Blackwell.
- Katz, L.F. (1986) *Efficiency Wage Theories: A Partial Evaluation.* Working Paper No 1906. Cambridge, Mass: NBER.
- Katz, L.F. (1988) "Some Recent Developments in Labor Economics and Their Implications for Macroeconomics", *Journal of Money, Credit, and Banking*, Volume 20, Number 3, pp. 507-522.
- Kendall, M.G. (1948) *Advanced Theory of Statistics* (Vol. II). London: Charles Griffin and Company.
- Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money.* London and Basingstoke: MacMillan.
- Kindleberger, C.P. (2000) *Manias, Panics, and Crashes. A History of Financial Crises.* 4<sup>th</sup> Edition. Wiley.
- Knut Wicksel, J.G. (1907) "Krisernas gåta", *Statsøkonomisk Tidsskrift*, Volume 21, pp. 255-286.
- Kydland, F.E. and E.C. Prescott. (1977) "Rules Rather Than Discretion: The Inconsistency of Optimal Plans", *Journal of Political Economy*, Volume 85, Number 3, pp. 473-491.
- Kydland, F.E. and E.C. Prescott. (1982) "Time to Build and Aggregate Fluctuations", *Econometrica*, Volume 50, Number 6, pp. 1345-1369.

- Laibman, D. (1997) *Capitalist Macrodynamics: A Systematic Introduction*. London: Macmillan Press.
- Lauderdale, J.M. (1804) *An Inquiry into the Nature and Origin of Public Wealth*. Second Edition. Edinburgh: Archibald Constable & Co.
- Lindbeck, A. (1992) "Macroeconomic Theory and the Labor Market", *European Economic Review*, Volume 36, Numbers 2-3, pp. 209-235.
- Lloyd, S.J. (1837) *Reflections Suggested by a Perusal of Mr. J. Horsley Palmers's Pamphlet*. London: P. Richardson.
- Long, J.B. and C.J. Plosser. (1983) "Real Business Cycles", *Journal of Political Economy*, Volume 91, pp. 39-69.
- Lucas, R.E. (1975) "An Equilibrium Model of the Business Cycle", *Journal of Political Economy* Volume 83, December.
- Lucas, R.E. (1976) "Econometric Policy Evaluation: A Critique", In K. Brunner & A. Meltzer (Editors), *The Phillips Curve and Labor Markets*. Amsterdam: North Holland.
- Malthus, T.R. (1820) *Principles of Political Economy*. London: John Murray.
- Mankiw, N.G. (1985) "Small Menu Costs and Large Business Cycles: A Macroeconomic Model", *Quarterly Journal of Economics*, Volume 100, Number 2, pp. 529-538.
- Mankiw, N.G. (2006) "The Macroeconomist as Scientist and Engineer", *Journal of Economic Perspectives*, Volume 20, Number 4, pp. 29-46.
- Marx, K. (1893) *Das Kapital*, Volume II. Berlin: Dietz, 1971.
- Matthews, R.C.O. (1959) *The Trade Cycle*. Cambridge: Cambridge University Press.
- Mills, J. (1867) "Credit Cycles and the Origin of Commercial Panics", *Transactions of the Manchester Society*.
- Mundell, R.A. (1963) "Capital Mobility and Stabilization Policy under Fixed and Flexible Exchange Rates", in R. MacDonald & M.P. Taylor (Editors), *Exchange Rate Economics*. Aldershot, U.K.: & Brookfield, US; Edward Elgar, 1992, pp. 50-60.
- Myrdal, G. (1939) "Fiscal Policy in the Business Cycle", *American Economic Review*, Volume 29, Number 1 (Supplement Papers and Proceedings), pp. 183-193.
- Nelson, C.R. and C.I. Plosser. (1982) "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications", *Journal of Monetary Economics*, Volume 29, Number 2, pp. 139-162.
- Parkin, M. (1986) "The Output-Inflation Trade-off When Prices Are Costly to Change", *Journal of Political Economy*, Volume 92, Number 1, pp. 200-224.
- Phelps, E.S. (1968) "Money-Wage Dynamics and Labor Market Equilibrium", *Journal of Political Economy*, Volume 74, pp. 678-711.
- Phillips, W. (1828) *Manual of Political Economy*. Boston: Hilliard, Gray, Little and Wilkins.
- Pigou, A.C. (1927) *Industrial Fluctuations*. London: MacMillan.
- Plosser, C.I. (1989) "Understanding Real Business Cycles", *Journal of Economic Perspectives*, Volume 3, Number 3, pp. 51-78.
- Prescott, E.C. (1986) "Theory Ahead of Business Cycle Measurement", *Federal Reserve Bank of Minneapolis Quarterly Review*, Volume 10, Number 4, pp. 9-22.
- Reijnders, J. (1997) "Tugan-Baranowsky's Breakthrough in Business Cycle Theory", in J. Glombowsky, A. Gronert and H.W. Plasmeijer (Editors), *Zur kontinentalen*

- Geschichte des ökonomischen Denkens*  
*History of Continental Economic Thought*.  
 Marburg: Metropolis Verlag, pp. 211-238.
- Romer, D. (2006) *Advanced Macroeconomics*. Third Edition. Boston: McGraw-Hill & Irwin.
- Samuelson, P.A. (1939) "On the Interaction between the Multiplier Analysis and the Principle of Acceleration", *Review of Economics and Statistics*, Volume 31, pp. 75-78.
- Sargent, T.J. and N. Wallace. (1975) "Rational Expectations, the Optimal Monetary Instrument and the Optimal Money Supply Rule", *Journal of Political Economy*, Volume 73, April.
- Sargent, T. J., & Wallace, N. (1976) "Rational Expectations and the Theory of Economic Policy", *Journal of Monetary Economics*, Volume 2, April, pp. 169-184.
- Schumpeter, J. (1927) "The Explanation of the Business Cycle", *Economica*, Volume 21, December, pp. 286-311.
- Schumpeter, J.A. (1939) *Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process*. New York: McGraw Hill.
- Sherman, H.J. and D.X. Kolk. (1996) *Business Cycles and Forecasting*. New York: HarperCollins College Publishers.
- Simonde de Sismondi, J.C.L. (1819) *Nouveaux principes d'économie politique*. Paris: Delaunay.
- Slutzky, E. (1927) "On the Summation of Random Causes as the Source of Cyclic Processes", *Problems of Economic Conditions*, Volume 3, Number 1.
- Slutzky, E. (1937) "On the Summation of Random Causes as the Source of Cyclic Processes", (extended version) *Econometrica*, Volume 5, Number 1, pp. 105-146.
- Snowdon, B. and H.R. Vane. (2005) *Modern Macroeconomics: Modern Macroeconomics. Its Origins, Development and Current State*. Cheltenham: Edward Elgar.
- Spiethoff, A. (1925) "Krisen", *Handwörterbuch der Staatswissenschaften*, Volume 6, Number 6, pp. 8-91.
- Stiglitz, J.E. (1999) "Toward a General Theory of Wage and Price Rigidities and Economic Fluctuations", *American Economic Review*, Volume 89, Number 2, pp. 75-80.
- Sweezy, P.M. (1942) *The Theory of Capitalist Development*. 11<sup>th</sup> Edition. New York: Monthly Review Press, 1970.
- Taylor, J.B. (1980) "Aggregate Dynamics and Staggered Contracts", *Journal of Political Economy*, Volume 88, Number 1, pp. 1-23.
- Tinbergen, J. (1936) "Kan hier te lande, al dan niet na overheidsingrijpen, een verbetering van de binnenlandse conjunctuur intreden, ook zonder verbetering van onze exportpositie?", In VVSS (Editor), *Prae-adviezen van de Vereeniging voor de Staathuishoudkunde en de Statistiek's*. Gravenhage: Martinus Nijhoff, pp. 62-108.
- Tobin, J. (1980) "Stabilization Policy Ten Years After", *Brookings Papers on Economic Activity*, pp. 19-89.
- Tobin, J. (1992) "Keynesian Theory: Is It Still a Useful Tool in the Economic Reality of Today?", in M. Baldassarri (Editor), *Keynes and the economic policies of the 1980s*. New York & London: St. Martin's Press/Macmillan Press in association with *Rivista di Politica Economica*, pp. 3-10.
- Tobin, J. (1996) *Full Employment and Growth: Further Keynesian Essays on*



*Policy*. Cheltenham, U.K. and Lyme, N.H.: Elgar; distributed by American International Distribution Corporation, Williston, Vt.

Tugan-Baranowsky, M. (1901) *Studien zur Theorie und Geschichte der Handelskrisen in England*. Jena.: Fisher.

Wade, J. (1833) *History of the Middle and Working Classes*. London: E. Wilson.

Weiss, A. (1990) *Efficiency Wages: Models of Unemployment, Layoffs, and Wage Dispersion*. Princeton: Princeton University Press.

Weiss, L. (1986) "Efficiency Wage Theories: A Partial Evaluation: Comment", in S. Fischer (editor), *NBER Macroeconomics Annual 1986*. Cambridge, Mass. & London: MIT Press, 285-287.

Wood, G. (1999) "Great Crashes in History: Have they Lessons for Today?", *Oxford Review of Economic Policy*, Volume 15, Number 3, pp. 98-109.

Woodford, M. (2003) *Interest and Prices*. Princeton & Oxford: Princeton University Press.

Yellen, J.L. (1984) "Efficiency Wage Models of Unemployment", *American Economic Review*, Volume 74, Number 2, pp. 200-205.

Yule, G.U. (1927) "On a Method of Investigating Periodicity in Disturbed Series with Special Reference to Wolfer's Sunspot Numbers", *Philosophical Transactions of the Royal Society of London* (A 226), pp. 267-298.

Zarembka, P. (2003) (Editor) *Value, Capitalist Dynamics, and Money*. Amsterdam: Reed Elsevier.

Zarnowitz, V. (1998) "Has the Business Cycle Been Abolished?", *Business Economics*, Volume 33, Number 4, pp. 39-45.

Zarnowitz, V. (1999) "Theory and History behind Business Cycles: Are the 1990s the Onset of a Golden Age?", *Journal of Economic Perspectives*, Volume 13, Number 2, pp. 69-90.

Jan Reijnders  
 Utrecht School of Economics  
 Netherlands  
 j.reijnders@uu.nl

## Business Cycles: Endogenous Heterodox Perspectives

*Howard J. Sherman*

### **Introduction**

This article examines the methodology of business cycle research, three theories of the cycle, the main relationships that must enter a complete model of the cycle, and some conclusions and policy implications.

Unemployment can not be analyzed at an individual level, but only at the level of the entire relations of society. Interviewing individual workers about their work preferences would be worthless for an understanding of the cycle – they all want to work, though some may be disheartened by a bleak outlook. Rather, one should begin by asking what kinds of tensions in the social-economic fabric may cause mass unemployment? At the present time, one must consider not only the institutions of one country, but world wide institutions.

Second, understanding the business cycle demands rejection of an equilibrium approach specifying the eternal laws of equilibrium, because such eternal laws do not exist. Rather there must be an investigation of the specific historical era in which we exist, that is, the institutions and dynamics of global capitalism. Pre-capitalist societies have no exchange, no use of money, and no profit maximization. Only modern capitalism has exchange for profit in the market and production for profit as its central features.

This approach, based on the institutions of capitalism, is spelled out in Burns and Mitchell (1946). They also spell out an empirical method of describing and testing different cycle theories.

Finally, one cannot assume that every individual has the same power, the same relations to the wealth of the world. Rather, the specific institutions of capitalism embody economic relations between different classes. To understand consumption, one must understand that most consumption comes from the income of workers and that worker's income is limited by the relationship of workers to their employers. Investment cannot be understood unless one recognizes that a relatively small class of people control most corporate assets and base their investments on the profit outlook. Moreover, relatively small classes of wealthy owners of corporations wield immense political power in each nation and in global capitalist institutions.

The best place to start the empirical study of the cycle under capitalist economic institutions is Burns and Mitchell (1946). For the Keynesian theory of instability, the best place to start is Keynes (1936). For the history of Marxist theories of the cycle, the best place to start is Sweezy (1942). For a comprehensive history of the classical and neoclassical approaches to the business cycle through the 1950s, see Haberler (1960).

### **Different Views**

The basic neoclassical view is that the economy is normally in equilibrium and downturns merely result from exogenous shocks. Thus, unemployment is the result of individual choices resulting from these shocks. This view is stated most clearly in Lucas (1975).

The best rebuttal of the neoclassical view is provided in the works of the institutionalist, Wesley Mitchell (1951). He proves in great detail that each stage of the business cycle is explained by, and only by,

the preceding stage. This complete endogenous explanation makes unnecessary any exogenous explanation, except as a minor additional factor.

Another view is the theory of fundamentalist Marxists that the rate of profit has a long-run tendency to fall. It falls because there is an increasing ratio of physical capital to labor, but only labor provides profit. For details, see Shaik (1978).

Critics argue that (1) the empirical assumptions of the theory are false, (2) the logic of the theory is incorrect, and (3) it is irrelevant to the business cycle because it is based on long-run trends.

Similarly, theories of long-run decline due to a continuous tendency of consumption to fall are also sharply critiqued. The critics say the long-run underconsumption theory cannot explain the cycle of boom and bust (see explanation and criticism in Bleaney 1976).

### **Underconsumption Theories**

There are three different endogenous theories of the short-run cycle. Each deals with the "Profit Squeeze." According to the profit squeeze view, in each expansion, falling profit eventually reduces investment and causes a contraction.

One theory may be called the short-run Underconsumption theory. It claims that in every capitalist expansion, the wages and salaries of employees do not rise as fast as the profits of employers. Since employees have a much higher consumption ratio (consumption/income), the shift of income to the corporate owners means a declining consumption ratio. The declining consumption ratio means that business cannot sell all its products to realize the potential profit embodied in those products

of labor. Declining profit causes a downturn. The best such theory is given in Sweezy (1942)

While this theory obviously reflects some aspects of the business cycle of global capitalism, critics have argued that its basic flaw is that it pays no attention to business costs. Profit, however, cannot be squeezed from the demand side alone. Profit is always revenue from demand minus the costs of supply.

### **Reserve Army Theory**

A second endogenous, short-run theory may be called the Reserve Army of Unemployed theory. It argues that wages of labor are held down at the beginning of the expansion by a large number of unemployed workers, the Reserve Army of the Unemployed. As the expansion progresses, however, the reserve army is depleted as more production adds new jobs. Eventually, after some time lag, this raises wages relative to production. This lowers the ratio of profit to total revenue, so it lowers the profit rate. The lower profit rate leads to less investment, which sets off a recession. This theory was first presented in a detailed and logical manner in Boddy and Crotty (1975).

According to critics, the basic problem with this theory is that it pays no attention to demand. Costs alone cannot cause a decline in the profit rate; profits are squeezed only if higher costs occur at the same time as limits on demand.

### **Nutcracker Theory of Profit Squeeze**

The third theory says that profits are squeezed from both sides, as in a nutcracker, so it has been dubbed the Nutcracker theory. It says that profit is squeezed in every expansion by rising costs and limited demand. It is explained in full in Sherman,

1991. It is also the framework for the following discussion of empirical relationships

Marx (1867), Keynes (1936), and Mitchell (1951) all agree on some major points of a profit squeeze theory. They do not conform to Say's law, but discuss a deficiency of effective, aggregate demand. They all agree that inequality of income affects aggregate demand. They all agree that interest costs and raw material costs rise rapidly in expansions. They all agree that profit is squeezed at the cycle peak by both limited consumer spending and rising costs of raw materials and interest. The declining profit leads to less investment, which causes a contraction.

The first mathematical model showing profit squeezed by both limited consumption and rising costs was Kalecki (1935). Also see Sherman (1991).

### **Empirical Relationships over the Cycle**

The empirical relations of the profit squeeze or nutcracker model are discussed in detail in Sherman (1991) and in Sherman (2003).

The usual econometric correlations and regressions do state important relationships, but do not tell us how each variable behaves over the cycle. The best way to begin to understand what actually happens in the cycle is to use the nine-stage description of the behavior of a variable in the method created by Wesley Mitchell (see discussion of Mitchell's method in Arthur Burns and Wesley Mitchell 1946, and discussion of it in Sherman 2001). Using Mitchell's descriptive method, as well as econometrics where useful, gives the following results.

Investment is the key variable determining expansion of output, jobs, and income. Investment always rises more rapidly than consumption in the expansion

and falls more rapidly in contractions. Investment is determined by profit expectations, which can be approximated by total profits. There is a time lag between investment and profits because of the need for businesses to collect the necessary information, plan new physical investments, borrow the money to make the investment, and spend the money over time. Profits commonly rise rapidly in expansions and fall rapidly in contractions. But profit leads production and investment at the cycle peak, while also leading at the cycle trough.

The question is how to explain this profit behavior. By definition, profits of each corporation equal revenue minus costs. The question is how to calculate these in the aggregate for a country and for the world. Following Keynes, the total revenue will equal consumer spending plus investment spending plus government spending. In addition each country will have net exports(positive or negative), though they cancel out for all of global capitalism. On the cost side, there is the costs of human labor plus the cost of material goods (raw materials, plant and equipment) and revenues to other agents (interest, rent, and taxes).

Before going on to discuss each variable, two points are necessary to know about the global picture. First, detailed data mostly relates to each country, but in general the sequence of happenings is the same in cycles all over the world and the relative magnitude of swings in each variable is also somewhat the same in each country.

The U.S. business cycle has been dominant for several decades in the sense that the other cycles tend to follow it. Therefore, Table 1 shows the cycle trough and peak dates for the U.S. cycle. Table 1 certainly reveals a cyclical movement, but it

has no significant periodicity. The cycles of the 1960s and 1990s were unusually long, while some were unusually short.

*Table 1*

TROUGHS	PEAKS
1949.4	1953.3
1954.2	1957.3
1958.2	1960.2
1961.1	1969.4
1970.4	1973.4
1975.1	1980.1
1980.3	1981.3
1982.4	1990.3
1991.2	2001.1
2001.4	2007.4

*Source:* Adapted from NBER (2009).

Note: EG, 1949.4 means fourth quarter 1949

Second, the data on the turning points of each country show much synchronization in the world—and some indication of increasing synchronization as the economic ties of world become closer and as the system of capitalism has become the sole dominant economic system in almost every country (see, e.g., the outstanding study by Arthur MacEwan 1984). Thus the following empirical statements certainly apply to the United States (with the largest single economy), but they also apply to most other countries and to the whole world to a large extent.

### **Important Variables**

One important variable is employee compensation (wages and salaries), both in absolute size and as a ratio to national or world income. In every cycle, compensation rises in the expansion and falls in the contraction. But the wage ratio—that is, the ratio of compensation to national income—falls in every expansion, while usually rising

in contractions. It is, however, a leading indicator, so it only falls in the expansion up to the peak of profits, then slowly rises to the cycle peak.

Similarly total consumption rises in expansion, largely because of rising employee compensation. But the consumption ratio to income falls during the expansion, following the path of the wage ratio. These facts explain part of the limitation of demand toward the end of expansion.

Government spending rises slowly in every expansion, but rises more rapidly in contractions because of rising expenditures on welfare, unemployment compensation, and business subsidies in areas such as agriculture. Taxes rise rapidly in every expansion of income, while falling rapidly in every contraction. As a result, there is an automatic tendency for deficits to fall during expansions and rise during contractions. Less deficit spending near the peak of expansion is another reason limiting demand.

In each country, imports rise as national income rises, but fall as national income falls. But exports follow world demand, which does not follow exactly the same cycle as any one country. Therefore exports tend to rise and fall more slowly in each country than its imports. As a result, next exports of each country tend to fall in expansions—another factor limiting demand—while net exports tend to rise in contractions, helping demand. These movement cancel out in the world as a whole. What is important for global capitalism is that the total volume of trade tends to fall in a world contraction because the downturn of trade in one country transmits to all other country an additional cause for decline.

One important factor of rising costs in an expansion in one country and in the world is the rise of raw material prices. Raw material prices tend to rise and fall faster than finished goods prices. Therefore, the cost of raw materials hurts the profits of industrial countries during the expansion. The same phenomena helps the profits of raw material producing countries in expansions, then hurts their profits very badly in each contraction. Since the business cycle always originates in the advanced, industrialized capitalist countries, the fact that rising raw material prices hurt profits in those countries means that a recession may begin, which then spreads to the less developed, raw material producers.

Another cost which hurts industrial capitalist profits is the rate of interest. The rate of interest generally rises during expansions, thus hurting profits towards the end of expansion. It falls during the contraction, helping profits to recover. But the rate of interest is a lagging indicator. It continues rising after the cycle peak because consumers and businesses are forced to borrow more money for a while to meet their unexpected lack of income. It also continues to fall for a while at the beginning of expansions because there is a large supply of unused funds, while people and corporations are still wary about borrowing.

### **Story of the Cycle**

There are no permanent crises, but some seem awfully long to those of us living through them. Eventually, however, there will be a recovery from the present stagnation or rolling recession.

#### *Expansion Scenario—Recovery*

The recovery is the first half of expansion in Mitchell's terminology. The outlook for

future profits has brightened and production rises. With rising production comes a remarkable rise in profits, the most rapid period of rise in profits. With a time lag, investment rises at its fastest rate of growth. Why do profits rise? On the demand side, rising production means rising employee compensation, almost all of which is spent for new consumer goods. Government grows slowly at this point, but does grow. Imports are still very low as people are still careful about luxury spending.

On the cost side, employee compensation rises more slowly than product, partly because workers are not yet accustomed to a rising economy, so are unwilling to strike. The falling compensation share is also due to the fact that overhead workers (such as guards and bookkeepers) need not be increased proportionately. A little of the downward pressure on compensation is due to rising technology reducing the need for employees.

Taxes rise at about the same rate as national income. Raw material prices rise very, very slowly because there are still big reserves of raw materials. Interest rates actually fall at the beginning of the recovery, only slowly starting to rise as the recovery continues.

Thus in the recovery, demand is rising rapidly, while costs are still quite low. As a result profits must rise, so investment must rise.

#### *Expansion Scenario—"Prosperity"*

Mitchell calls the last half of expansion "prosperity." In reality, it may be a slow rise or a run away boom in the economy and in the stock market. Also, in reality, the labor share drops, so the income of the poor, the middle-income workers, and even professional workers falls far short of the

increase in product. Those employers with a low wage bill are very happy. But employers are also sellers of goods, so they should not be so happy. If compensation does not rise as rapidly as product, demand falls below supply. An employer should be truly happy only if his wages are kept low, but all other employers are paying higher wages.

In addition to the falling labor compensation ratio, demand is also hurt because government spending rises more slowly than government taxes. This means that there is a decline in deficit spending and often an actual rise in surplus taken by taxes. The surplus means that government may spend less on interest, but it also means a decline in the amount of net government demand for goods.

At the same time, in a country with an expanding economy, imports rise rapidly with national income rising, but exports usually do not rise as rapidly because other economies are not fully synchronized. This means less net exports, that is, less demand for the goods of that country.

The declining labor compensation ratio – which lowers the growth of consumption – plus the decline of net government spending, plus the decline in net foreign demand for exports, all mean that demand is limited as the prosperity continues. At the same time, some costs are rising faster than production is rising.

Demand for raw materials is a function of the change in demand for finished goods. Therefore, in prosperity there is a rapid rise in the prices of raw materials. As raw material prices rise faster than finished goods prices, in the materials-using nations, who produce finished goods, the cost of raw materials is one factor tending to lower profits.

Another rising cost is the rate of interest. As the economy heats up and people become very optimistic, consumers borrow more money on the assumption of rising compensation, while corporations borrow more money on the assumption of rising sales. The optimistic demand for loans meets a finite supply of money for loans. Therefore, interest rates begin to rise, cutting into profits.

Profits are thus squeezed from both directions. On the one side, consumer demand is slowing its rate of growth, net government demand flowing into the economy is declining, and net export demand is declining in most of the advanced capitalist countries. As demand declines in one country, it means lower exports for other countries. At the same time, costs of raw materials and interest rates are rising. Of course, taxes are rising at the same rate as national income, while employee compensation is also rising, though at a slower pace than production or national income. Since profits are menaced by both limited demand and rising costs, profits fall in many countries. When profits fall, there will eventually be a fall in investment. The declining investment in several countries means less new jobs, which means lower employee compensation, which means an actual decline in consumer demand (though credit may keep consumption high for a while). The decline in several large countries leads to a decline in global capitalism, since trade is falling for more and more countries, profit is falling, and new investment is falling.

#### *Contraction Scenario—Crisis*

Mitchell calls the first half of contraction the crisis. Just as there is a rising circle of affluence in recovery, there is a vicious

circle of decline in the crisis. To spell it out, in country after country demand is limited by falling or stagnant consumer demand, by decline in net government spending, and by decline in net exports. At the same time, raw material prices are very high, while interest rates actually continue to climb in early contraction. As more and more countries go into recession, the disease is transmitted to the whole world. Instead of other countries forming a market for domestic surpluses, their demand is also declining. Moreover, the pessimism in one country spreads to the investors of other countries. In the age of immediate information, it is the usual rule to see that all stock markets rise together in a boom, but all stock markets fall together in a bust, with even daily movements often mimicked in every market around the world.

#### *Contraction Scenario—Depression*

Mitchell calls the last half of contraction the depression. At first the depression grows worse and worse with the spiral of falling jobs and rising unemployment, less consumer demand, and less investment. But eventually all of these tendencies reverse themselves – after great human suffering.

Total employee compensation falls. The labor share of income, however, rises because profits fall more rapidly than compensation. The rising labor share, even from a declining national income, supplies a floor to consumer spending, so the consumption ratio begins to rise—even though total consumption is falling.

The government spending rises to cover unemployment compensation, welfare, and business subsidies, all automatic under present law in many countries. But taxes are falling. Therefore, the government deficit rises rapidly, providing a net flow of money into the economy. This aids the recovery of

demand. Similarly, in many countries imports usually fall faster than exports, so net exports may increase in the depths of depression.

On the cost side, interest rates fall in the depression. The low cost of borrowing helps improve the outlook. Raw material prices also decline, often precipitously. This hurts demand in the less-developed countries who export natural resources, but it helps the profit picture in the industrialized countries.

At the bottom of the depression, therefore, the future begins to look better for profits. Demand has reached a floor, while costs are still declining. This leads to new investment and a new expansion.

#### **Conclusions and Policy Implications**

The institutions of capitalism tend to produce a cycle of boom and bust. Of course, the cycle may also be affected by external shocks, such as floods or wars or terrorism. These usually do not change the basic economic direction, but may reduce the momentum or exacerbate the boom or the bust. The boom is caused by the floors to demand and the continuing fall of costs at the bottom of the depression. The bust is caused by the ceiling to demand and the rising costs toward the end of prosperity. Several kinds of remedies have been tried.

Monetary policy can reduce an inflationary expansion by tightening credit, though too much tightening can bring on a recession. In a severe contraction, even very low interest rates have little effect on increasing credit, so they are little help for demand in such circumstances.

Fiscal policy can use tax cuts for low income people as well as increases in government spending on welfare and unemployment compensation to stimulate demand. If this is done on a massive enough



scale, it can end a depression and cause an expansion. But such a massive boost occurs only during a massive war effort, such as the Second World War—which ended the Great Depression.

But massive deficits cause problems. Interest on the national debt is greatly increased, which causes a large shift of income from the average taxpayer to the very rich (who hold most of the bonds). Moreover, the way to stop inflation is the opposite: reducing government spending and increasing taxes. Therefore, if there is a combination of unemployment and inflation—which has occurred at some times in the late twentieth century—there is no fiscal cure; and certainly no monetary cure.

To fight unemployment and inflation at the same time, one policy that works in the short-run is expansionary monetary and fiscal policy combined with price controls. Price controls worked to some degree to prevent inflation in the face of enormous deficit spending in the Second World War (with the use of physical rationing ), but there was repressed inflationary pressure, which showed itself when controls were lifted. Price controls under Nixon held back spending by severe limitations on wages and salaries, with mild controls on prices. But any regime of real price controls combined with stimulation can only work in the short-run. Eventually, it distorts resource allocation as demand and supply change, the political manipulation adds so much uncertainty that investment and innovation are reduced, and the political strength of corporations ensures that income distribution becomes more unequal. Long-run price controls are not compatible with capitalism.

Another method of control is central planning and extensive government ownership. There is no example of this

under democratic political institutions — though France had extensive public ownership for a short while and some planning, which proved ineffective because France is too small a part of the international market in which it is integrated. The Soviet Union was a case of dictatorship and central ownership and planning of almost everything. It did prove that this non-capitalist system could produce full employment at all times and rapid growth for a while. It also proved that the combination of dictatorship and central ownership of everything is highly inefficient, blocks innovation, and is subject to repression of freedom.

The case of France shows that the only effective way to control the boom and bust cycle of global capitalism is control at the global level. That means trying a democratic world government, global ownership of the largest corporations, and democratic planning of those corporations to end the cycle of boom and bust. Theoretically, this should work, but there is no proof one way or the other.

### **Selected References**

- Bleaney, Michael. (1976) *Underconsumption Theories*. New York: International Publishers.
- Boddy, Raford, and James Crotty. (1975) "Class Conflict and Macro-Policy." *Review of Radical Political Economics* 7 (Spring): 1-17.
- Burns, Arthur, and Wesley Mitchell. (1946) *Measuring Business Cycles*. New York: National Bureau of Economic Research.
- MacEwan, Arthur. (1984) "Interdependence and Instability: Do the Levels of Output in the Advanced Capitalist Countries Increasingly Move Up and Down

Together?" *Review of Radical Political Economics* 16: 57-79.

Haberler, Gottfried. (1960) *Prosperity and Depression*, 4th ed. Cambridge, Mass.: Harvard University Press.

Lucas, Robert. (1975) "An Equilibrium Model of the Business Cycle." *Journal of Political Economy* 83 (Dec.): 1113-44.

Keynes, John M. (1936) *The General Theory of Employment, Interest and Money*. New York: Harcourt Brace Jovanovich.

Kalecki, Michal. (1935) *Theory of Economic Dynamics*. New York: Monthly Review Press, 1968.

Marx, Karl. (1867) *Capital*. Volume 1. Chicago: Charles Kerr, 1903.

Mitchell, Wesley. (1951) *What Happens During Business Cycles*. New York: National Bureau of Economic Research.

NBER. (National Bureau of Economic Research) (2009) *US Business Cycle Expansions and Contractions*. New York: NBER. [www.nber.org/cycles/cyclesmain.html](http://www.nber.org/cycles/cyclesmain.html)

Shaikh, Anwar. (1978) "An Introduction to the History of Crisis Theories", in Union for Radical Political Economics (URPE) (ed.), *U.S. Capitalism in Crisis*, pp. 219-48. New York: Monthly Review Press.

Sherman, Howard. (1991) *The Business Cycle: Growth and Crisis in Capitalism*. Princeton, N.J.: Princeton University Press.

Sherman, Howard and David X. Kolk. (1996) *Business Cycles and Forecasting*. New York: Harper Collins.

Sherman, Howard. (2001) "The Business Cycle Theory of Wesley Mitchell," *Journal of Economic Issues*, Volume 35, Number 1, March, pp. 85-98.

Sherman, Howard. (2003) "Institutions and the Business Cycle," *Journal of Economic Issues*, Volume 37, Number 3, Sept. 2003, pp. 621-642.

Sweezy, Paul. (1942) *Theory of Capitalist Development*. New York: Monthly Review Press (reprinted 1970).

Howard Sherman  
 Department of Political Science  
 University of California  
 Los Angeles, USA  
[sherman@polisci.ucla.edu](mailto:sherman@polisci.ucla.edu)

## Central Banking

*Chris Moore*

### **Introduction**

From humble beginnings in the seventeenth century, central banks have risen to be at the apex of the financial institutional structure of national economies. They are pre-eminent in the conduct of monetary policy and play a dominant role in ensuring financial stability. Increasingly, the functions of central banks are taking on an international perspective, as a result of the growing integration of financial markets. The Bank for International Settlements (BIS) lists 132 central banks and currency boards on its website.

The evolution of central banks as key players in economic affairs has a colourful history embodying over two centuries of economic and political thought. Their status and functions has been subject to the tides of fashions in economic and political philosophies. This is as true today as it has ever been. Even though central banks are increasingly adopting similar practice there is still vigorous debate about: the structure of central banks and related governance issues; the nature of monetary policy objective(s) and how they should be implemented to best achieve economic objectives such as stability of prices and output. Other lively debates, with practical significance to the effectiveness of central banking revolve around the degree of central bank independence and issues relating to accountability and transparency.

Major policy and operational functions of central banks include: the design and implementation of monetary policy; the supervision of the banking and payments systems to ensure their smooth operation;

banker and fiscal agent to the government e.g. manager of public debt; lender of last resort to other financial institutions; issuer of currency—notes (paper money) and coins; and the promotion of the overall stability of the financial system, especially in times of financial crisis. Nevertheless, not all central banks carry out all of these functions. While all central banks have responsibility for monetary policy implementation, for example for historical, practical or theoretical reasons, the supervision of banks may be carried out by a separate supervisory agency. Not only do the functions of central banks vary from country to country but central banks are also evolving institutions. As Deane and Pringle (1995:2) aptly put it: “...despite the distinguished pedigrees some boast, a central bank as we know it today is a twentieth century phenomenon – a Johnny-come-lately. So unlike its ancestors is it, that it must be classed as a entirely distinct species”.

### **Historical Perspective**

The first central bank, the Bank of the Estates of the Realm was established in Sweden in 1668—renamed the Riksbank in 1866. It started life as a private bank, the Stockholm Banco in 1656, but failed due to issuing too many notes without sufficient collateral. The Swedish parliament, the Riksdag, then took over its management. Shortly after, in 1694 the Bank of England was founded and became the Government’s banker and debt-manager. In doing so, it brought much needed order to England’s hitherto chaotic system of money and credit. In the early days, the creation of central banks meant that a nation’s financial resources could be efficiently mobilised to fund wars and national development. According to Siklos (2002), the motivation

for establishing both the Riksbank and the Bank of England, and indeed the 1782 forerunner of the Bank of Spain, was to finance war. These three were the only central banks up until 1800, when the Banque of France was established.

The nineteenth century, was a fertile one with eleven central banks established. The principal ones being the Reichsbank, the forerunner of the Bundesbank in 1876, the Bank of Japan in 1882, and the Banca d'Italia in 1893. The motivation for establishing these banks was no longer the explicit funding of military ambitions but a combination of bringing order after a financial or economic crisis, the management of public debt, profits from seignorage (the profit arising from the difference between the face value of notes and coins and their actual cost of production), the promotion of economic growth and the creation of a single note issuing authority.

The majority of central banks in the nineteenth century were in Europe. The rest of the world meanwhile experimented with a variety of forms of monetary authority, the most popular being the currency board – a monetary authority issuing notes and coins convertible on demand to a foreign currency, the reserve currency, at a fixed rate and effectively therefore operating on a gold standard monetary theory. The British established the first currency board in the colony of Mauritius in 1849. After some years of trial and error the currency board form became orthodox with the creation of the West African Currency Board in 1912. By the 1930s they were common throughout the British colonies of Africa, Caribbean, Asia and the Pacific Islands. There is still interest in the currency board concept amongst economists as they are seen as a

viable alternative to central banks. Currently there are twelve currency boards operating including those in Hong Kong, Gibraltar and Brunei.

The twentieth century witnessed rapid growth in central bank numbers. The most significant central bank formed in the early years of the century was the American Federal Reserve System (Fed) in 1913. Some eighty five years later the European Central Bank (ECB) established in 1998, took responsibility for the monetary policy of the euro area in 1999. The formation of the ECB represents a landmark in the history of central banking since national central banks have surrendered their autonomy in the conduct of monetary policy to a regional institution. Between the establishment of the Fed and the ECB, more than one hundred new central banks arrived on the scene worldwide, mainly as a result of ex-colonies creating their own central banks from currency boards on gaining independence.

In the last twenty years the profile of central banks has risen considerably as monetary policy has taken over from fiscal policy, as the principal tool of economic management. With this transition central banks have moved on from being lenders of last resort to assuming the role of guardians of price stability. This has placed central banks, and their governors, firmly in the driver's seat of economic policy-making.

## **Main Functions of Central Banks**

### *Monetary Policy*

The development and conduct of monetary policy is arguably the primary function of a modern central bank. Monetary policy, is currently the primary macroeconomic tool, although once playing second fiddle to fiscal policy, especially in the post-Second World

War period when Keynesian discretionary stabilisation policy was the orthodoxy. Today, the high profile of central bank governors and their message of low inflation growth are well known to the public. Their monetary policy announcements and economic commentaries make headlines and move markets. Monetary policy outcomes impact household budgets directly by affecting return on savings, the cost of debt, including housing mortgages, the prices of goods and services and the purchasing power of savings.

Over the past two decades central banks with the help of theoretical advances in economics and in particular public choice thinking (e.g. Nordhaus 1975), have reshaped the conduct of monetary policy, with success in the monetary policy area being signalled by the advent of low inflation growth in most developed economies. Mishkin (2000a) outlines several guiding principles that underpin success in monetary policy. Although generally accepted, these principles are not incontrovertible truths and are still debated.

The first principal is that price stability, or low and stable inflation, benefits the economy by increasing the level of productive resources employed. However, studies of the link between inflation and economic growth have not conclusively shown that low inflation is always associated with higher levels of economic growth. Nevertheless, it is generally accepted that high inflation levels are not conducive to economic growth. This is important, for in many countries price stability is the sole objective of monetary policy. Inflation targets are often stringently specified. In New Zealand for example, a Price Targets Agreement is negotiated between the Governor of the Reserve Bank and the

Government's Minister of Finance/Treasurer and sets the price stability band - currently 1–3%, after allowing for specified price shocks.

It is noteworthy that monetary policy is as concerned to avoid deflation (a fall in the general price level) as it is with inflation. Deflation can severely undermine the stability of the banking system and economic performance. If deflation leads to a precipitous decline in asset prices e.g. of equities and property, a loss of wealth effect could result in lower consumption spending and a reduction in the level of collateralised bank lending. An inability of banks to lend because of reduced collateral, or of equity markets to raise capital, will stifle investment and economic growth. If deflation got out of hand there is potential for major bank failures as assets become worth less than the loans secured against them.

The second principle is monetary and fiscal policy alignment. It is easier to achieve price stability if monetary policy does not have to contend with an inflationary fiscal policy. If in opposition, monetary policy needs to be tighter in order to achieve price stability, so imposing unnecessary economic cost. The other risk is that with large fiscal deficits there is the temptation to pressure the central bank to monetise the debt (print money), thus increasing money supply and inflation.

The third principle is to avoid the time inconsistency problem, (see Kydland and Prescott 1977), which arises when a central bank attempts to take advantage of the short-run trade-off between inflation and employment in order to boost employment in the short-term. Expansionary monetary policy will in the long-run lead to higher inflation (as inflationary expectations get

built into wages and prices) and subsequent job losses in the future, so negating the short-term job effect. Although the problem is well known, central banks may come under political pressure to boost jobs in the short-term. The time inconsistency problem is therefore a major reason for a general move to central bank independence and the setting of a single price stability goal for monetary policy.

The fourth principle is the need for monetary policy to be forward-looking. The time lag between implementation and effect—from 3 to 24 months, before monetary policy actions start to influence the variables that underpin inflation e.g. wage setting behaviour or output changes; makes it essential that monetary policy be forward looking. If left too late, with response to merely current inflationary trends, inflation expectations would have taken hold and become embedded in pricing and wage setting behaviour. The costs of reining in the inflationary horse once it has bolted can be high as then the monetary policy stance has to be more restrictive than would have been required had it been pre-emptive rather than reactive. Of course a forward-looking monetary policy needs to be well informed. Hence, central banks invest heavily in sophisticated economic modelling and forecasting exercises. Without a clear understanding of how the economy works, or a knowledge of dynamic structure of the transmission mechanism - how monetary policy actions influence economic and monetary variables, it is difficult to recognise the signs of incipient inflationary pressure sufficiently early so pre-emptive monetary policy action can be taken.

The fifth principle is accountability. This is application of the adage ‘no responsibility without accountability’, or ‘responsibility

plus accountability equals results’. In other words, governors who do not achieve specified monetary policy objectives, should be sanctioned or removed. Indeed a number of central bank governors are on performance-based contracts where contract renewal is based on meeting clearly stated monetary policy objectives.

The sixth principle states that the ultimate objective of monetary is a healthy economy not just price stability per se. Monetary policy should also be concerned with output and seek to minimize output fluctuations within the context of price stability. It is fair to say that even where monetary policy objectives are stated in terms of price stability, policy actions are rarely taken without consideration of its impact on output.

The final guiding principle states that serious economic downturns are most often associated with financial instability. The corollary is that it is important for central banks to promote financial stability if it is going to achieve its goal of low inflationary economic growth.

Bringing the seven principles together to define a central bank’s institutional features for conducting monetary policy operations Mishkin identifies that price stability should be the overarching long-term goal of monetary policy. Further, an explicit nominal anchor should be adopted such as a fixed exchange rate, or money-growth target, or as New Zealand and others like Canada the UK have done, an explicit inflation target may be stated. Such explicit anchors help overcome the time inconsistency problem and assist in aligning fiscal policy. A central bank should on the one hand be goal dependent, which means an elected government or policymakers set the monetary policy goal. If the public are

unhappy with the goal they can change the government. This democratic approach assists in overcoming the time inconsistency and fiscal alignment problems. On the other hand the central bank should be instrument independent. That is once the government has mandated the monetary policy goal then the central bank should be free to decide the way (what instruments) it will use to achieve the goal. Independence in this sense helps avoid political interference to exploit the inflation-employment trade off and allows central banks to be forward looking in the conduct of monetary policy. Central banks should also be accountable to government and the public. The adoption of a nominal anchor can help here by offering an explicit measure of its performance for all to see. The legal relationship the central bank has with government must allow for some measure of censure to occur i.e. through the Governor's contract, or by way of some legislative sanction.

Monetary policy was once a mysterious affair conducted in smoke-filled rooms. This is far from the case today. Transparency of actions and clear communications of monetary goals and actions now form an essential aspect of a successful monetary policy. A better-informed and knowledgeable public are more likely to buy-in to price stability if they understand its benefits and how the central bank intends to achieve it. Hence central banks now make publicly available (in hardcopy and on websites) a host of economic and financial statistics, research papers, economic forecasts and speeches of Governors and other senior bank management. It is argued that an economically informed public helps reduce political pressure on the central bank, as well as restrain fiscal policy. Openness also assists businesses and households plan

by reducing uncertainty over the future path of inflation (prices and wages), interest rates, and foreign exchange rates.

### *Instruments of Monetary Policy*

The chain of events linking central bank monetary policy operations to inflation outcomes is collectively known as the transmission mechanism. It describes how central bank monetary operations - which generally target overnight interest rates by way of targeting a quantity (to influence its price), or the price directly, of a specified bank asset or reserve (the instrument)—influences the price of other financial variables e.g. longer-term interest rates, the exchange rate, and how these variables in turn influence expenditure and investment decisions that ultimately determine the level of inflation.

Central banks have a diverse array of tools and instruments in their armoury with which to implement monetary policy. Their choice of instrument(s) will depend on the institutional arrangements defining central bank operations and relationships with the banking system, the varied nature and sophistication of money markets across jurisdictions, the types of traded securities available, economic efficiency objectives and historical practice. There is nevertheless a common theme running through the way central banks implement monetary policy. In order to implement monetary policy a central bank has leverage over either the quantity or the price of an asset.

The procedures used by the Reserve Bank of New Zealand make for a good exemplar because it recently moved from a quantity target regime to an explicit interest rate (price) setting one. Under the old quantity approach, the asset or reserve was called primary liquidity—it comprised

banks' settlement cash balances held at the Reserve Bank and assets, including Reserve Bank bills, exchangeable on demand for settlement cash. The leverage the Reserve Bank had over registered banks was that they must use settlement cash (they still have to today) to settle interbank obligations that arise from the millions of daily transactions that take place between bank customers, including those of the government. By setting a quantity target for settlement cash the Reserve Bank was able, in an indirect way, to influence the level of overnight interest rates as banks competitively bought and sold settlement cash to meet their requirements. A lowering of the settlement cash target signalled a firming or tightening in monetary policy as its reduced quantity was expected to lead to rise in its interest rate (price). Conversely, an increase in the target signalled a loosening of monetary policy and an expected fall in interest rates.

In common with all central banks, the Reserve Bank of New Zealand conducted (it still does) open market operations (OMOs) in order to remove or add liquidity, from or to, the banking system to maintain the settlement cash target. The conduct of OMOs is required to offset known net cash flows that result from government revenue flows associated with tax and expenditures e.g. welfare payments, and the banking system. The process of conducting OMOs involves making detailed forecasts of government's net cash flows several months ahead. The Reserve Bank, like all central banks, uses a variety of tools with evocative names like repos (repurchase agreements) to inject cash into the system for a specified period, or inject cash outright by purchasing government securities from banks' liquidity portfolios. In order to withdraw cash from

the banking system, central banks typically sell government securities to banks. Under New Zealand's old quantity regime banks also held special purpose Reserve Bank bills that could be converted to settlement cash via a discount window. A discount window, common to many central banks, allows a bank in need of liquidity to convert specified financial securities at a discount i.e. a discount margin is added to its yield so reducing its price. The discount level applied may reflect the stance of monetary policy and can vary with the maturity profile of the security being discounted.

New Zealand moved away from a quantity to a price regime because of the unpredictable relationship between the quantity of settlement cash and overall monetary conditions (Archer *et al.* 1999). To overcome the deficiencies in the quantity regime, the Reserve Bank became reliant on the use of market statements to signal its view of the appropriateness of monetary conditions—irreverently referred to as open mouth operations—OMOs” of a different kind! After much discussion and debate, the Reserve Bank adopted, along with most central banks, an implementation strategy based on setting the price of settlement cash. The Official Cash Rate (OCR) is set every six weeks. The Reserve Bank pays 25 basis points (0.25 percentage points) below the OCR on money deposited in settlement accounts overnight, and provides cash by way of repos at 25 basis points above the OCR. Overnight interest rates are expected will lie within the 50 basis wide ‘corridor’ set by the interest rates at which the Reserve Bank will inject and adsorb overnight funds on demand. The adoption of the OCR saw the end of Reserve Bank bills and discounting, but open market operations remain integral to managing government



cash flows that impact on the level of settlement cash. The OCR regime's 'corridor' approach is also used in Australia and Canada.

An alternative approach to supplying cash on demand as the Reserve Bank does, is that of discretionary operations as practised in the USA. Here the market is influenced towards the target federal funds rate, by the Federal Reserve System injecting or withdrawing liquidity. Kasman (1992) compares and reviews the monetary policy operating procedures in the USA, Germany, Japan, the UK, Canada and Switzerland. The review finds that intervention strategies are similar in the countries studied in that they use interest rate operating objectives, and implement policy by way of OMOs – supplying and absorbing reserves at market prices – albeit using different financial instruments

### *Prudential Supervision*

Supervision of banks, or more generally of the financial system, is an important role of a central bank. However, today the supervision role is not exclusive to central banks. A number of countries have separated the monetary policy role from that of bank supervision by handing over supervision to an independent institution. In 1997, for example, Gordon Brown, the then UK Chancellor of the Exchequer, took the bank supervisory duties from the Bank of England and gave it to the UK Securities and Investment Board. (See later sub-section, *Independence*, for further discussion).

No matter who is responsible for supervision, it is paramount that the stability of the financial system is managed so that a bank failure is contained and possible contagion (knock-on) effects are minimised. If a bank failure is allowed to spread to

sound banks then the viability of the whole financial system, including the payments system, could be at risk. The economic costs of a systemic failure are enormous. Many of the most severe economic downturns in history have their root in the collapse of the financial system.

There are many approaches to bank supervision. They range from the hands-off bank disclosure regime unique to New Zealand, to very hands-on, supervisory regimes fully armed with bank inspectors and supported by bank deposit insurance, as found in the USA. No one approach is inherently superior to another. Over recent years prudential supervision has, however, moved away from the "regulatory approach" to that of a "supervisory approach" (Mishkin 2000b) with less emphasis now placed on bank compliance with regulations e.g. quality of banks' balance sheets and capital adequacy, though still important. Instead the supervisory focus has shifted to examining the soundness of banks' management practices—risk management systems, as a means of controlling risk. This change reflects the increased complexity of the banking environment associated with the rapid innovation in markets and trading instruments that allow employees to build up huge losses in short time periods. The Barings failure of 1995 is a classic example of poor management practice.

A comprehensive 107 cross-country study by Barth *et al.* (forthcoming) of the relationship between bank regulations and supervisory regimes, and bank development, performance and stability, found that supervisory regimes that encourage accurate information disclosure, the private or corporate ownership of banks (as opposed to government ownership) and private monitoring of banks, create the best

environments for bank development, performance and stability. In particular, the study found that restrictions on bank activities (e.g. into insurance, real estate etc) were not positively correlated with bank development or stability. This finding supports the view that the ability to diversify income sources generally enhances bank stability. It also found that restrictions on the entry of foreign-owned banks, in particular the specific impediments, are positively associated with bank fragility. The study also found, after controlling for bank regulations and supervision that stringent capital regulations did not lead to greater bank development or stability. Interestingly, generous deposit insurance schemes were found to be associated with bank instability, and were not able to be mitigated with strong official supervision or capital standards. Finally, the study failed to find a robust relationship between the level of supervisory power, independence and provisioning stringency, and bank development, performance and stability. The findings suggest that an effective bank supervisory regime, one that promotes bank development, performance and stability, should comprise regulations and supervisory practices that force accurate information disclosure, encourages the private ownership of banks, private monitoring (market discipline) and bank entry (new participants), and reduces moral hazard. The findings are in sympathy with the three “Pillars” approach adopted in the new Basel Capital Accord II, especially its third pillar, with its emphasis on the role of market discipline (see De Ceuster & Masschelein 2003 for a discussion on market discipline. See also section *Other Tasks* for a brief discussion of Basel II).

The traditional approach to safeguarding the financial system is for the central bank to act as lender of last resort i.e. supply liquidity to a bank or the financial system that is short of money to meet obligations to customers, including other banks. The central bank’s objective in supplying liquidity is to stop the financial crisis spreading. The problem bank may or may not be insolvent i.e. its liabilities exceed its assets. If the bank is insolvent then the expectation is that the bank will be ultimately eased out of the financial system unless it receives an injection of new capital.

The existence of a safety net such as lender of last resort can give rise to moral hazard. It can create an incentive for a bank to take on excessive risk in the knowledge that it will be rescued if it fails, so leading to greater rather than less financial instability. To lessen the risk of moral hazard it is important that the financial sector knows that the central bank will only exercise the role in exceptional circumstances.

A challenge for supervisors is the “too big to fail” problem. This arises when governments are reluctant to let large banks fail because depositors and other stakeholders are likely to suffer substantial losses. It may also lead to systemic failure and trigger a financial crisis. Bailing out a large bank tends to exacerbate moral hazard and raises the question of how best to limit moral hazard in large financial institutions. One solution involves adopting the supervisory approach—concentrating on sound management practices, and in the use of the market discipline as an incentive to less risky behaviour.

#### *Other Tasks*

Central banks carry out a number of other important tasks. One is the design,

production and distribution of notes and coins in circulation. This may be done directly by the central bank or on its behalf by another agency. For example, in the USA note production is done by the Treasury Department's Bureau of Engraving and Printing under the direction of the Board of Governors. The notes are purchased at cost not face value. In Europe the production of the euro is carried out in each member state with overall control lying with the European Central Bank (ECB).

Central banks may or may not be directly involved in determining the nature of foreign exchange (FX) policy i.e. fixed or floating. For example, in the USA and Europe, neither the Fed nor the ECB determine FX policy although they would be consulted over any proposed change in the FX regime because of its centrality to the conduct of monetary policy. Lack of consultation would severely jeopardise price stability.

Central banks may also service government by acting as the government's fiscal agent, and holding and managing official reserves. The fiscal agent role is often at arm's length; as the central bank is likely to be prohibited from purchasing government debt (printing money) to fund deficits, although it can, of course, manage government bond issues (to finance fiscal deficits) on behalf of government.

Overseeing the efficient operation of the payments system—fundamental to both the stability of the financial system and the conduct of monetary policy—is an important task for central banks in most countries. In some cases central banks like the Fed and the ECB operate clearing and payment systems, as well as oversee private providers of payments services. Included here is the enforcement of prohibitions on cross border transactions that contravene UN sanctions or

transactions suspected of being terrorist or money laundering related.

Other central bank activities may, depending on the jurisdiction, include the promulgation of financial regulations and legislation to protect consumers, particularly in the area of credit. Central banks also have a role in the collection and publication of financial statistics and to undertake research for public dissemination. This, as already mentioned, plays a part in enhancing central bank transparency and furthers its creditability in both the formation and operation of monetary policy.

Finally, on the global stage, central banks work together through a number of cooperative endeavours and organizations such as the Bank for International Settlements that aims to limit the risks inherent in international transactions. Supervision and general bank risk management is one area of substantial central bank cooperation through the Basel Committee on Banking Supervision. Though not attempting to harmonise banking supervision in detail, the Basel Accord, first adopted in 1988, lays out broad supervisory standards, guidelines and best practice for central banks to adopt. The hope of such accords - Basel Core Principles and the International Capital Accord - is that a global convergence of standards and approach to bank supervision and the management and control of banking risks, at a general level, will occur. Most recently a comprehensive review of the Accord has resulted in Basel II—a new Accord is proposed, based on three "Pillars." The first sets out minimum capital requirements, the second focuses on standards for official supervisory review of capital adequacy and the third involves financial market

discipline, to complement the supervisory review.

## **Issues Facing Central Banks**

### *Governance and Governing Structures*

The nature and structure of central bank ownership is of interest because of the actual or perceived level of political influence it may afford the government over the conduct of monetary policy. Central banks in many countries are state-owned institutions and operate under legislation similar to that of other government bodies. The high level of state ownership dates back to the aftermath of the great Depression and a time in the late 1940s when governments believed that the state had a major hands-on role to play in the running of the economy. Of the twenty-one banks cited in Siklos (2002) seventeen are state-owned, including the ECB, four (Austria, Belgium, Italy and Switzerland) are a mixture of state and other, including public and private shareholders. The Fed is an exception in that banks in their respective region own the twelve regional banks comprising the Federal Reserve System; although ownership is not strictly a commercial arrangement.

Relevant also to the issue of ownership and political interference is the appointment of central bank Governors or Presidents and governance. Siklos (2002) argues that the concern over appointments of Governors is misplaced for two reasons. First, bank personalities only come into play at times of a financial crisis and do not generally impact on routine day to day operations of central banks. Second, central banks have responded to the challenge of meeting changing government and societal goals, by becoming more transparent and accountable in the way they communicate decisions

concerning the implementation of monetary policy. The personality and style of central bank Governors or Presidents is thus less important today than in the past. Walsh (2002) maintains that dismissal rules can help ensure accountability.

Central banking can be a very profitable business and therefore ownership can reap considerable economic rent. Profits from functions and activities of central banks, such as seigniorage, open market operations, and fees, are distributed in accordance with ownership structure. In the case of state-owned central banks, profits go to the state based on a formula or budget, or to the Finance Minister or the Treasury. This is even true for the Fed, for despite the appearance of private ownership, on average 95% of the Fed's considerable profit, goes to the US Treasury.

Central banks have a long history of adaptation that continues today with vigorous debate over fundamental issues like, the tasks central banks should perform, the level of independence they ought to enjoy, how accountability can be best achieved, and the related issue of openness and transparency, especially in the conduct of monetary policy. The debate is essentially concerned with finding optimal governance and operational structures that allow a central bank to most efficiently achieve its mandated goals (e.g. price stability, stable financial system) within a democratic society. The debate is well informed with new economic thinking and academic writings. These issues are also close to politicians' hearts, as economic management has traditionally been in the political domain. Perceived loss of economic control to unelected central bankers is politically less desirable. Nevertheless, governments have, and are making, bold decisions to give

central banks more autonomy in the interests of enhanced economic performance. But with such power comes demands for accountability, especially as central banks seek greater public credibility by becoming independent and hence perceived to be less susceptible to political influence.

### *Independence*

Freedom to decide how to pursue its goals and difficulty to reverse its decisions by political authorities, are the two defining aspects of central bank independence (Blinder 1998). A number of academic studies over the past two decades (see e.g. Berger *et al.* 2001) have shown that the greater the degree of independence the lower the country's inflation rate. The theoretical explanations for this inflationary bias are the time-inconsistency problem discussed above and/or the revenue motives arising from inflation tax. One approach to overcome the inflationary bias problem is to make central banks independent and thus free from political interference. This has been the approach taken by the many governments including the Fed, the ECB and countries like the UK, Canada, New Zealand and Australia. The theories underpinning central bank independence are the conservative-central-banker approach where over time the central banker develops a reputation for maintaining price stability but at a cost of increased output variability (Rogoff 1985), although Alesina and Summers (1993) did not find greater variability of employment and output associated with independent central banks in OECD countries. The other theory is the principal-agent approach, where a contract imposes a cost on the central banker when inflation moves away from the optimal (see e.g. Walsh 1995). In both cases the central bank needs to be

independent from political or societal pressures so that it can act to achieve and maintain low inflation in an optimal (least economic cost) way. As central banks have become more independent so have the calls for greater accountability to government and the public. This has generally meant a more open, communicative and transparent *modus operandi* than was traditionally the case, particularly in the formulation and conduct of monetary policy. Issing *et al.* (2001:128) however, go so far as to assert that: "In democratic societies, independence goes hand in hand with accountability".

### *Separation of Tasks*

Debate continues over the wisdom of separating monetary policy and supervisory/regulatory roles with arguments for and against. It can be argued as do Peek *et al.* (1999) that the two roles are complementary. They found that the information gathered in the process of bank supervision can improve the accuracy of central banks' macroeconomic forecasts of important variables like employment, output, inflation, all of which help fine-tune the settings of monetary policy instruments that lead to superior monetary policy performance. It might also be argued that any informational gaps arising from separating the roles could be managed by an informational swapping arrangement.

The conflict of interest argument is a major reason given for functional separation (Goodhart 1995). An example of such conflict is when higher interest rates are necessary for the central bank to dampen inflation yet this might adversely impact on bad debts and the health of the financial system. There is a trend to separate the tasks of bank supervision and monetary policy by taking the supervisory task away from the

central banks and placing it in a different institution. The rationale for this division of labour is the presumption that the supervisory and monetary policy tasks are not necessarily compatible. Goodhart, however, maintains that a better solution to conflict problems might be “greater statutory clarification of objectives, rather than institutional separation” (1995:344). It might also be pointed out that the first duty of the central bank is to oversee the stability of the financial system rather than aid monetary policy and therefore it should be carried out by an institution best placed to do it. Another reason put forward for central banks to retain bank supervision is that it is also the lender of last resort in many jurisdictions. To be effective in this role it needs an intimate knowledge of the failing institution which it is unlikely to have if supervision is carried out by another agency.

## Conclusion

Towards the end of the twentieth century Stanley Fischer (1996) wrote that the enduring challenge facing central banks is to fight inflation. Even though a low inflation climate in the developed countries appears to be cemented in, the principal challenge for central banks in the twenty-first century will be to keep inflation low and financial systems safe. For despite the advent of central bank independence and performance contracts, the conduct of monetary policy and bank supervision will not be easy tasks in an environment of rapidly innovating financial markets and instruments, that include the emergence of non-bank e-money, e-banking and e-finance services. The ability to store money in electronic form, for example, could hasten the disintermediation of banks as depository and payment institutions. Depository banks play

a critical role in monetary policy’s transmission of interest rates. Hence the adoption of e-money could pose a serious threat to the continuance of low inflation and financial stability. It has even been suggested that the existence of non-bank e-money could make central banks obsolescent in the future (Stevens 2002). It will be interesting to see how governments and their central banks respond to this challenge. The ECB, for example, already has regulations in place limiting the provision of e-money to institutions subject to reserve requirements. How long such regulations can limit the impact of future financial innovations on central bank core functions is a moot point.

## Internet Sites

Asian Development Bank [www.adb.org](http://www.adb.org)  
Bank for International Settlements  
[www.bis.org](http://www.bis.org)  
International Monetary Fund [www.imf.org](http://www.imf.org)  
World Bank [www.worldbank.org](http://www.worldbank.org)

## Selected References

- Alesina, A. and L.H. Summers. (1993) “Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence”, *Journal of Money, Credit and Banking*, Volume 25, Number 2, pp. 151-162.
- Archer, D., A. Brooks and M. Reddell. (1999) “A Cash Rate System For Implementing Monetary Policy”, *Reserve Bank of New Zealand: Bulletin*, Volume 62, Number 1, pp. 51-61.
- Barth, J.R., G. Caprio Jr. and R. Levine. (Forthcoming) ‘Bank Regulation and Supervision: What Works Best?’, *Journal of Financial Intermediation*.

- Berger, H., J. de Haan and S. Eijffinger. (2001) "Central Bank Independence: An Update of Theory and Evidence", *Journal of Economic Surveys*, Volume 15, Number 1, pp. 3-40.
- Blinder, A.S. (1998) *Central Banking in Theory and Practice* MIT Press, Cambridge, Mass.
- Deane, M. and R. Pringle. (1995) *The Central Banks* Viking Penguin, New York.
- De Ceuster, M.J.K. and N. Masschelein. (2003) "Regulating Banks through Market Discipline: A Survey of the Issues", *Journal of Economic Surveys*, Volume 17, Number 5, pp. 749-766.
- Fischer, S. (1996) "Central Banking: The Challenges Ahead – Maintaining Price Stability", *Finance and Development*, Volume 33, Number 4, pp. 34-37
- Goodhart, C.A.E. (1995) *The Central Bank and the Financial System* MIT Press, Cambridge, Mass.
- Issing, O., V. Gaspar., I. Angeloni and O. Tristani. (2001) *Monetary Policy in the Euro Area* Cambridge University Press, Cambridge.
- Kasman, B. (1992) "A Comparison of Monetary Policy Operating Procedures in Six Industrial Countries", *Federal Reserve Bank of New York Quarterly Review*, Summer, pp. 5-24.
- Kydland, F. and Prescott, E. (1977) "Rules Rather than Discretion: The Inconsistency of Optimal Plans", *Journal of Political Economy*, June, pp. 473-91.
- Mishkin, F.S. (2000a) "What Should Central Banks Do?", *Federal Reserve Bank of St. Louis Review*, November-December.
- Mishkin, F.S. (2000b) "Prudential Supervision; Why is it Important and What Are the Issues?" National Bureau of Economic Research, *Working Paper*, Number 7926.
- Nordhaus, W.D. (1975) "The Political Business Cycle", *Review of Economic Studies*, Volume 42, pp. 169-190.
- Peek, J., E. Rosengren and G. Tootell. (1999) "Is Bank Supervision Central to Central Banking?" *The Quarterly Journal of Economics*, Volume 114, Number 2, pp. 629-53.
- Rogoff, K. (1985) "The Optimal Degree of Commitment to an Intermediate Monetary Target", *Quarterly Journal of Economics*, Volume 100, Number 4, pp.1169-90
- Siklos, P.L. (2002) *The Changing Face of Central Banking* Cambridge University Press, Cambridge.
- Stevens, E. (2002) "Electronic Money and the Future of Central Banks", *Economic Commentary*, Federal Reserve Bank of Cleveland, March.
- Walsh, C.E. (1995) "Optimal Contracts for Central Bankers", *American Economic Review*, Volume 85, Number 1, pp.150-67.
- Walsh, C.E. (2002) "When Should Central Bankers be Fired?", *Economics of Governance*, Volume 3, pp. 1-21.

Chris Moore

Department of Finance, Banking & Property

Massey University

New Zealand

C.I.Moore@massey.ac.nz

## Chartalism and Metallism

*Sergio Rossi and Szymon Klimaszewski*

### **Introduction**

Chartalism and metallism are two schools of thought that have been confronting since the inception of economic analysis. Their aim is to explain the origins, nature, and value of money in a logical and historical framework, but according to different, and in many ways opposing, paradigms.

Although the origins of both schools may be traced to the work of Plato and Aristotle respectively, the expressions *chartalism* and *metallism* were first used by Knapp in 1905 only (Knapp 1924). In spite of their relatively recent coinage, however, both tags have been, and are, used to identify different points of view at different points in time. We thus need to clarify the contents and evolution of each line of thought, before any attempt at policy analysis on the grounds of either paradigm.

### **Metallists versus Chartalists**

Initially, the debates between the adherents to metallism and the advocates of chartalism centred on money's value: metallists viewed it as the factual result of the value of money's stuff, that is, a commodity, whilst chartalists argued that it stemmed from social trust and convention. This debate led quite naturally to expand on the origins and nature of money.

#### *Metallism: Money is a Creature of Markets*

The metallist point of view can be traced back to the sixteenth century (see Goodhart 1998, Bell 2001) and led to some debates between the banking and currency schools in the eighteenth and nineteenth centuries. According to metallism, money is logically

as well as historically a commodity, whose value derives from the stuff that makes it up, independently of its monetary role in economic transactions.

This vision has been elaborated upon by Menger (1892), who, in fact, puts forward an explanation of how a particular commodity has gradually evolved to become a medium of exchange, owing to a spontaneous market selection process driven by a transactions cost minimizing behaviour of rational agents. More recently, Menger's approach was taken up by a number of neoclassical economists, led by authors such as Brunner and Meltzer (1971), Ostroy (1973), Jones (1976), Alchian (1977), and Kiyotaki and Wright (1989, 1991, 1993), the latter two authors elaborating a 'search theory' of money's origin that is now the mainstream approach to monetary economics.

In a nutshell, these authors consider money as "a creature of the market", in the sense that it was generated by a search process that agents spontaneously carried out to solve the practical problem of the so-called "double coincidence of wants" existing in barter trade. "Think, indeed, of the peculiar difficulties obstructing the immediate barter of goods in those cases, where supply and demand do not quantitatively coincide; where, *e.g.*, an indivisible commodity is to be exchanged for a variety of goods in the possession of different persons" (Menger 1892:242).

To minimize transactions costs, traders were led to discover that commodities have what Menger (1892:242) dubs "*different degrees of saleableness*". As commodities are more or less saleable in respect of the greater or less ease with which they can be disposed of at any convenient time and at current market prices, according to Menger (1892:244–5) traders are led by the market



mechanism of supply and demand to identify a commodity generally accepted in exchange for all sorts of real goods and services. This commodity, which came to be identified with different items in different historical periods, then becomes a medium of exchange in the Friedman sense that “[it] enables the act of purchase to be separated from the act of sale” (Friedman 1974:8).

The definition of money that stems from this approach considers therefore that “any commodity to be called ‘money’ must be *generally acceptable in exchange*, and any commodity generally acceptable in exchange should be called money” (Fisher 1911:2). Hence, although money may wield other functions, it is thought of as having originated as a medium of exchange, in a variety of forms like rocks, leather, furs, spice, tobacco, slaves, and wives (sic!); and more recently in the form of precious metals—gold and, to a lesser extent, silver—owing to their intrinsic or physical properties (most importantly their homogeneity, divisibility, portability, and durability; see Clower 1967, Spindt 1985). On account of these intrinsic characteristics pertaining to the medium of exchange, metallists argued that the advent of paper money further reduced transactions costs as well as those market “frictions” that money helps “lubricate” (on frictions in monetary economics, see Niehans 1978:16).

In developing this market-based view, and on account of the benefits induced (at least in theory) by perfect competition on the market for produced goods and services, some supporters of metallism went as far as adhering to the free banking view, that is, a competitive money supply that would give rise, “democratically”, to a social consensus on the best (to wit, inflation-proof) medium of exchange (Klein 1974, Hayek 1978,

Selgin and White 1994, Dowd 2000). This led metallists to elaborate their view in order to provide an answer to the crucial question as to why “every economic unit in a nation should be ready to exchange his goods for little metal disks apparently useless as such, or for documents representing the latter” (Menger 1892:239). The metallists’ answer was that the value of paper money derives from the intrinsic value of its metal backing (see Menger 1923), so much so that, in a number of countries, and for monetary policy practices, paper money has been *de jure*, but often also *de facto*, convertible into a precious metal (usually gold) at a fixed rate for a long historical period.

A further answer that metallists provided in order for their “real” theory to explain the value of a valueless ticket that agents use to enter into “the great social store of all goods [that are exchanged against it]” (Schumpeter 1954:289), argued that paper money is the “general equivalent”, or a “representative sign”, of all real goods and services on sale. As Ingham (1996:513) and Bell (2001:153) noted, this argument was based on Walras’s (1874) general equilibrium analysis, where money is the *numéraire* against which all other non-money goods are exchanged. According to Walras’s (1874:188) definition of the *numéraire*, however, “[o]ur standard of measure must be a certain quantity of a given commodity”, a definition that includes money in the set of commodities and raises therefore the Ricardo (1817) logically impossible problem of finding an invariant measure of value among the set of commodities existing within an economy (see Cencini 2001:13–15).

#### *Chartalism: Money is a Creature of Law*

Chartalists challenge the metallists’ view on a number of points, and indeed used to

consider themselves as anti-metallists, in the sense that they had not a positive theory to oppose to metallism originally (see Knapp 1924, and Schumpeter 1954).

Chartalists consider metallism as being too narrow, if not wrong, on both theoretical and historical grounds. In particular, Schumpeter (1954) focused on the logical origin of money and put aside its historical origins, while Knapp (1924) aimed at explaining money's validity in respect of institutions and laws. An ardent advocate of chartalism summarized the theoretical as well as the historical critiques to metallism, arguing that "the use of money does not necessarily imply the physical presence of a metallic currency, nor even the existence of a metallic standard of value. [...] [T]here is overwhelming evidence that there never was a monetary unit which depended on the value of a coin or on a weight of a metal; [...] in fact, there never was such a thing as a metallic standard of value" (Innes 1913:379). Indeed, the proponents of chartalism—or cartalism (Goodhart 1998) if one refers to the Italian word *carta* rather than the Latin word *charta* (both words meaning "paper" literally)—argue that money's value is, and has always been, independent of its material support (be it in a metallic or paper form).

Simmel (1907) was aware of this point and thus called for a distinction between the essence of money and the material used to wield its functions. In this respect, according to Goodhart (1989:34), "[t]he substitution of fiat, paper money, for metallic coin as the main component of currency in the last 200 years provides strong support for the Cartalist view that the monetary essence of currency can rest upon the power of the issuer and not upon the intrinsic value of the object so used." To put it differently, in the

chartalist view money is a creature of law, so much so that it signifies a (property) right sanctioned by a political authority—the state in modern societies (see Knapp 1924:39–40, but also Heinsohn and Steiger 1983, 1994). The state (or the "polis" in Aristotle's sense) plays an essential role in money's origins and value. In particular, according to Grierson (1977:33), money's origins "are not to be sought in the market but in a much earlier stage in communal development, when worth and wergeld were interchangeable terms."

As a matter of fact, the German term *Wergeld* means "penalty" and designates money to be paid for a murder or, generally speaking, for injuries inflicted to human beings in a given community. As noticed by Wray (1998:71), "[t]hese *wergeld* payments appear to be the source of some of our terminology. For example, the verb 'to pay' comes from *payer* and *pacare*, 'to pacify' or 'to make pace with' [...]. The word 'worth' comes from *Wert*, which when combined with *Geld* denotes the idea of measuring wealth and seems to have come from the practice of paying 'bride price' or 'bride wealth' compensation to a household for the loss of a daughter to marriage."

As a result, money originated as a unit of account and in the chartalists' view preceded market exchanges, which, as they argue, are a much later phenomenon in history (Polanyi 1977:123). According indeed to a number of historians, anthropologists, and sociologists, the market is only one possible "form of integration" of individuals in a community or society (see Polanyi 1944, Grierson 1977, and Ingham 2000). As Zazzaro (2003:228) points out, "[r]eciprocity—a form of socially obligatory donation—and redistribution—the assignment of individual or group production to the authority of the community

and the subsequent sharing out of goods to members of the community according to customs in force—are equally important, widespread social forms of integration, in which money may still perform its functions as a means of payment, unit of account and/or medium of exchange.”

In fact, even in ancient, stateless societies human relations were hierarchical and communitarian. They implied a unit of account in order to measure and regulate the reciprocity of obligations as well as the redistribution of goods; they also implied a means of payment in order for individuals to settle their social debts, such as those arising from status, kinship, convention, or religion (see Malinowski 1921, Einzig 1966, Polanyi 1977). In primitive societies money did not appear therefore as a market-induced result of a discovery process that costs minimizing agents went through to avoid the drawbacks of the “double coincidence of wants”.

The argument put forward, and elaborated upon, by chartalists is then that the origin of money is to be found in a political (sovereign) act that establishes by law, or by social convention, what object(s) people may dispose of in order for them to settle their debt obligations (Keynes 1930, Lerner 1947, Wray 1998). Thus, money’s value did (and does) not stem from its material support, be it a metal or paper object, but is based “on the antiquity of the law of debt” (Innes 1913:391). In fact, what Innes (1913:393) calls “the primitive law of commerce” is the basic principle of double-entry bookkeeping, which records debts and credits for further reference and settlement. From a technological viewpoint, this means that “money is equivalent to a primitive form of memory” (Kocherlakota 1998:232).

Debt–credit relationships, and records, have neither logically nor historically to do

with a particular material support. In other words, money’s value has no link with the stuff that carries out money’s function in both ancient and modern societies. It is society, or the state as argued by modern chartalists, that lies at the heart of it. The argument centres here on a social agreement, or on the political power of the state to impose that payments defined in a particular unit of account are made to the latter by the population. As Smith (1776:328) puts it, “[a] prince, who should enact that a certain proportion of his taxes should be paid in a paper money of a certain kind, might thereby give a certain value to this paper money” (see also Innes 1913:398–9).

By its willingness to accept a given paper money in the settlement of taxes and other debts, the state induces all taxpayers to accept these pieces of paper as money, because non-bank agents know for sure that everyone who has to pay taxes will accept them in turn (Tobin and Golub 1998:27). This argument can be found in the now widespread overlapping-generations approach to explain money’s existence and functions (see Wallace 1980, Geanakoplos 1987, Woodford 1990, and Handa 2000). But it also led some late twentieth-century economists to put to the fore a so-called taxes-drive-money approach (or “state theory of money”), by which the state is given a prominent place in the creation, circulation, and validation of money.

In general, the so-called modern neo-chartalists (Goodhart 1998, Wray 1998, Bell 2001) explain that sophistication of the penal system (wergeld, fines, and compensations), and a variety of increasingly complex fiscal tools (like taxes, duties, and fees), created a framework that compelled the state to fix a standardized unit of account for recording and finally settling obligations. In their view,

this explains the appearance of a unit of account defined by the state for the payment of any kinds of debt obligations by the administered population. These economists consider it as a relevant proof the fact that many accounting books were found in a number of centres of power, like palaces and temples.

As a matter of fact, rulers kept track of their outlays by different means, particularly by using clay, copper, or hazelwood tallies that were broken down in two parts, one part handed out to the supplier (who became a creditor) and the other to the buyer (who thus became a debtor) (see Innes 1913, Goodhart 1998, Wray 1998, 2003). To be sure, coins are just one possible form of token-money. They have been largely used as means of payment owing to the historical fact that the state has often a monopoly on the mints, and that with its *imprimatur* the state provides a universal guarantee against counterfeiting. However, a chartalist would never argue that the value of a coin is driven by its metal content, weight, or backing. As Innes (1913:382) pointed out long ago, “the monetary standard was a thing entirely apart from the weight of the coins or the material of which they were composed. These varied constantly, while the money unit remained the same for centuries.”

### **Theoretical Implications**

With the advent of bank (book-entry) money and the world-wide development of modern banking and settlement systems (domestic as well as cross-border), both metallists and chartalists were led to elaborate their views further.

At the theoretical level, metallists argued that bank money, which they acknowledged as being a double-entry in a bank's ledger, is the historical result of the practice by which

goldsmiths used to keep track of the (gold) commodities deposited with them for safe-keeping, in exchange for notes. “Eventually, the safe-keepers discovered the ‘deposit expansion process’. [...] In this case, the gold could remain within the safe, with a multiple number of notes circulating – both those initially created when depositors stored their gold but also all those created in loans” (Wray 2003:90). This is the now so-called “cloakroom theory of banking” *à la* Cannan (1921), and is subsumed under the causal chain by which deposits make loans, as is so often pointed out by those writers considering the financial intermediation role of banks (see Pesek & Saving 1968; King & Plosser 1984; but also Realfonzo 1998:chs 3–4 for a detailed historical account).

Chartalists, in this connection, argue that the metallist view is wrong, so much so in a monetary economy of production, where the loans that banks grant to firms in order for the latter to carry out production give rise to bank deposits by wage earners, a main tenet of the so-called French circuit school (see Graziani 2003). Further, as pointed out by Ingham (2000:23), “[b]arter exchange of commodities, whatever the complexity of the system, is essentially bilateral; but, monetary relations are trilateral.” In a monetary economy, as a matter of fact, “[e]very transaction involves three parties, buyer, seller, and banker” (Hicks 1967:11). The peculiar role of what in modern language is called a “bank” is to issue the means of final payment between the buyer (who is the payer) and the seller (that is, the payee) (on the notion of payment finality see Kahn and Roberds 2002, and Committee on Payment and Settlement Systems 2003). In a sociologist's words (Ingham 1996), money is a social relation; in an economist's jargon (Hicks 1975), money is a numerical counter,

which, moreover, is supplied by the banking system on the public's demand.

This led a number of chartalists, like Innes (1913:392, 402), to claim that money and credit are one and the same thing, a contention maintained also by Minsky (1986:228), who stated that "everyone can create money; the problem is to get it accepted". The chartalist argument revolves here around the acceptability idea, on which the modern version of this school builds a multi-tier pyramid to represent the hierarchy of modern monies (Bell 2001). The tiers of this pyramid represent different kinds of debt-credit relationships, each tier having its own degree of acceptability in a given economic system. The most generally acceptable form of money would be at the top of the pyramid (Foley 1989:250).

From this particular viewpoint, credibility, or trust, would be the necessary and sufficient condition for money's social acceptability, and circulation, as a means of final payment (see Dow and Smithin 1999). According to a number of neo-chartalists (Goodhart 1998, Wray 1998, 2003, Bell 2001, Bell and Nell 2003), this once again puts the state at centre-stage. More than by enacting legal-tender laws, as argued by Schumpeter (1954) and more recently by Davidson (1972), the state establishes the unit of account as well as the means of payment most widely used in an economic system, by determining what will be accepted "at its pay-offices, whether or not it is declared legal tender between citizens" (Keynes 1930:6, fn.1; see also Innes 1913:398; Knapp 1924:40; Lerner 1947:313).

In fact, the chartalist view of money's value based on acceptability is neither dependent on the state "writing the dictionary" (Keynes 1930:4) nor on the state

willingness to accept the established money in payment of taxes and other debt obligations. Chartalism is indeed essentially based on the society's need of trust and conventions (or even laws) to express, measure, record, and eventually settle individual-to-community (that is, part-whole) relationships (Ingham 1996, 2000). Taxation powers, fiscal policy, and a proper state are not necessary conditions for the chartalist theory about the origins, nature, and value of money. This school of monetary thought, in fact, allows for the existence of a variety of pay societies gravitating around a settlement institution – in the form of a clearing agent, which seems to have originated as a great periodical fair, where traders cleared their debts and credits without the use of a single coin (Innes 1913:396–7). As a matter of fact, this settlement institution represents the cornerstone of any modern network of debt obligations that may exist in the real world, independently of political powers.

On factual grounds, nevertheless, the state is certainly the most important agent in a pay society, hence its historical importance to explain the origins of money, owing to its extended and durable taxation powers on the administered population. This may increase money's dissemination within any economic system, but on logical as well as theoretical grounds is neither a sufficient nor a necessary condition to explain money's value.

### **Policy Implications**

According to metallists, in our monetary systems the central bank can and indeed does control the money stock, hence inflation, via the so-called money multiplier, which relates central bank money (or the money base, M0, often also dubbed high-

powered money by advocates of the money multiplier approach) to the targeted monetary aggregate (M1, M2, M3, M4, and so on). In short, the latter aggregate depends on the money base as well as on the actual value of a money multiplier,  $m$ , which in turn depends on the reserve requirements imposed by the central bank as well as on expenditure behaviour of non-bank agents with respect to the business cycle (see Handa 2000:Ch.10).

Positing that the relevant money multiplier is both predictable (by monetary authorities) and stable over time, modern metallists like Friedman (1974, 1987) and McCallum (1987, 1990) argue that the central bank can control the value of the money stock, hence inflation. The metallist argument revolves here around the functional need for banks to record central bank money in their ledger (reserves) in order for them to lend money to the public and thus create deposits in the process. By considering money's value as a logical result of its metal content, or backing, physically stored in the central bank's vault, metallists are led to the conclusion that the reserve ratio decided by the monetary authorities is crucial in order for bank money to preserve its value, therefore credibility, as time goes by.

However, the advent and wide diffusion of book-entry money, and later on the upsurge of various financial innovations—like asset and/or liability management, which have been put into practice by banks to avoid the reserve requirements imposed on them by the central bank—raised the multiplier stability problem with a renewed concern (see Goodhart 1989). In fact, in light of the relatively high liquidity degree of other forms of money than central bank's, banks and non-bank

agents have been switching back and forth between traditional bank deposits (demand or sight deposits) and new forms of bank deposits (time and saving deposits, deposit certificates, securitization, e-money products, and so on), moved by the interest rate differential and its variations over time. As a result the money multiplier became unstable and, moreover, unpredictable. This is so much so that it is of no practical relevance anymore in monetary policy making, because the ratio between the money base and the monetary aggregate targeted by the central bank shifts incessantly, making any monetary targeting strategy totally inappropriate for the control of inflation.

On account of the modest results that these monetary policy strategies led to in a number of countries in the 1980s and 1990s, namely as regards price level stability, neo-chartalists are in favour of a completely different policy. Following Lerner's (1943) functional finance approach, they call for a policy mix in which fiscal policy is expenditure-led and monetary policy interventions depend as a general rule on the contingent situation with respect to inflation as well as unemployment figures. Their view is based here on the argument that the state's expenditures give rise to fiat money, which the private sector then uses as a final means of payment owing to the state willingness to accept it back, and therefore to destroy it, in the payment of taxes. Hence Lerner's (1947:314) policy-oriented conclusion that "[t]he government—which is what the state means in practice—by virtue of its power to create or destroy money by fiat and its power to take money away from people by taxation, is in a position to keep the rate of spending in the economy at the level required to fill its two great responsibilities,

the prevention of depression, and the maintenance of the value of money.”

Neo-chartalists are thus utterly against the imposition of fiscal rules (expenditure caps, a balanced-budget requirement, or a deficit-to-GDP ratio such as the relevant criterion ruling a country’s membership to the European monetary union). These regulations tie a government’s hands in deciding the amount (and the nature) of public expenditures, with no misgiving for the macroeconomic situation of the country as a whole. They also lead to a pro-cyclical fiscal policy: increases in public deficits are only possible, and tolerated, when the economy is in an expansionary phase of the business cycle, but are severely limited, if not prohibited, during economic downturns, that is, when the private sector mostly needs them to boost aggregate demand and growth, hence to reduce unemployment.

In this respect, according to neo-chartalists the government has “the responsibility for maintaining the total rate of spending on goods and services at the level necessary to purchase all of the output that a fully employed labor force could produce” (Bell 2003:162; see also Wray 1998, 2003). These authors thus call for a series of monetary and fiscal policy interventions that would support output growth, particularly during a business cycle downturn or an economic crisis, when unemployment is high and actual output lies beyond its potential, full-employment level.

Neo-chartalists propose therefore that the government puts into practice what they call an “employer of last resort” (or labour “buffer stock”) programme. While there are different formulations of this programme (inspired by the Lerner functional finance view), all would have the government offer a productive job to anyone ready, willing, and

able to work, at a wage rate fixed by the government, and that any employer would consider as a benchmark against which he would fix his own wage scale and structure (see Harvey 1989, Gordon 1997, Wray 1998, 2003). “The government would essentially stand ready to ‘buy’ or ‘sell’ labor, offering jobs to any workers who showed up, or offering workers to any employers willing to hire workers out of the buffer stock” (Wray 2003:106). In an economic recovery or boom, this buffer stock would shrink, contributing thus to limit wage increases, because of the fixed wage floor that the state defines by holding constant the wage rate it offers. During a recession, by contrast, buffer stock employment would rise, thereby preventing wages from falling below the rate that the government defines in its programme.

By causing the government’s budget to move counter-cyclically, an employer of last resort programme would have an economic stabilization property: it would help to reach and preserve full employment as well as price level stability, since “the government is essentially offering to provide HPM [high-powered money] in exchange for labor” (Wray 2003:106). In a sense, such a programme increases produced output and employment, without generating inflationary pressures, because the remuneration of the public employment so provided generates the necessary and sufficient purchasing power, in the economy as a whole, to allow for market clearing at the prevailing market prices.

Moving from a national to an international monetary framework, neo-chartalists discard another key building block of the metallists’ theoretical construction, that is, the Optimum-Currency-Area (OCA) model put forward by Mundell

(1961) and taken up by several writers like McKinnon (1963), Kenen (1969), Fleming (1971), and all those advocating a loss of sovereignty in favour of monetary union (De Grauwe 2001). The OCA argument stems from the metallist idea that money is a medium of exchange, which, moreover, originated from a market mechanism by which agents aimed at minimizing transaction costs. If so, there is no economic reason for confining the use of a particular medium of exchange to a given political jurisdiction, or nation. In this framework, the sole criterion in forming a currency area is a cost-benefit analysis, to determine the geo-political boundaries within which it is optimal to have a single currency.

Comparing the benefits, mainly in respect of transactions costs minimization, of having a single currency in a wider area against the costs—in terms of lost adjustment forces once exchange rates have been fixed—would lead then to abandoning as a general rule the “one nation—one currency” equation. This moving away from the national definition of monetary areas is particularly promoted by advocates of the “less state, more market” view. From this point of view, as Wray (2003:92) observes, “many orthodox economists have applauded the creation of currency boards in (mainly) less developed countries on the argument that abandonment of monetary sovereignty by explicitly tying a nation’s (weak) currency to another nation’s (strong) currency helps to discipline profligate governments.” In short, a currency board would help importing stability and particularly with respect to inflation.

This argument, nevertheless, can be, and indeed has been, advocated by neo-chartalists as well, as noted by Wray (2003:107), in the form of dollarization, that

is, the adoption of a national currency (the dollar or the euro, for instance) by different countries as parallel, or unique, legal tender within their own geo-political boundaries (see Rochon & Seccareccia 2003). In this respect, in fact, neo-chartalists may also be criticized for putting too much emphasis on the state’s role when explaining the value of money, even if none can deny that modern money is essentially a symbol, or a representative sign of produced goods and services. The critique would be based here on the historical proof that the key currencies (such as the British pound and the American dollar, among a few others) “were initially accepted internationally, [but] not as a result of their acceptability by states in payment of taxes” (Rochon and Vernengo 2003:64). The reasons for their international acceptance stemmed from the financial role played by the banking systems issuing these currencies, in particular their providing “the bridge between the present and the future”, as so clearly put by Keynes (1936:293). As a matter of fact, the state cannot obtain commodities and services before the latter are produced, which implies firms asking banks for credit lines, and then paying out wages before government spending (see Rochon and Vernengo 2003). In short, production as well as bank money is essential to explain the origins, nature, and value of money. In this respect, the state and its (taxation) powers are epiphenomena.

## **Conclusion**

Chartalist and metallist monetary conceptions and policy implications are largely different. Yet there are also common grounds, which today are more pronounced than used to be the case. Metallists now consider money to be book-entry money, that is, an incorporeal final means of



payment whose value depends on produced output, against which money is exchanged. Neo-chartalists, for their part, view modern money as a unit of account that the state defines, and issues, to exert a property right over output to be produced by wage earners. These two views, which incidentally are by no means exclusive, converge therefore on the fact that money is the alter ego of produced output, or, still more precisely, that money is indeed the numerical container whose freight is given by those real goods and services that result from labour within a monetary economy of production.

To be sure, even if the state can induce, or force, the private economy to use money by a political decision, it can never determine the purchasing power of money independently of economic activity. As noted by the “founding father” of chartalism, money is the numerical definition of the very output it is associated to as in Cannan’s cloakroom theory of banking: “When we give up our coats in the cloakroom of a theatre, we receive a tin disc of a given size bearing a sign, perhaps a number. There is nothing more on it, but this ticket or mark has legal significance; it is a proof that I am entitled to demand the return of my coat” (Knapp 1924:31). Accordingly, the ticket, to wit, money, is the numerical form of the coat, that is, output, deposited in the cloakroom (a bank). The ticket holder has in fact a drawing right over the object deposited. Owing to the fungibility of bank deposits, it is obvious that deposit holders have a purchasing power over any real goods or services on sale within the corresponding monetary system. Thus, money and banking are the backbone of any payment activity between any two agents. All the rest, including the state, is ancillary to that.

## Selected References

- Alchian, Armen A. (1977) “Why Money?”, *Journal of Money, Credit, and Banking*, Volume 9, Number 1, pp. 133–40.
- Bell, Stephanie A. (2001) “The Role of the State and the Hierarchy of Money”, *Cambridge Journal of Economics*, Volume 25, Number 2, pp. 149–63.
- Bell, Stephanie A. (2003) “Neglected Costs of Monetary Union: The Loss of Sovereignty in the Sphere of Public Policy”, in Stephanie A. Bell and Edward J. Nell (eds) *The State, the Market and the Euro: Chartalism versus Metallism in the Theory of Money*, Cheltenham and Northampton: Edward Elgar, pp. 160–83.
- Bell, Stephanie A. and Edward J. Nell. (Editors) (2003) *The State, the Market and the Euro: Chartalism versus Metallism in the Theory of Money*. Cheltenham and Northampton: Edward Elgar.
- Brunner, Karl and Allan H. Meltzer. (1971) “The Uses of Money: Money in the Theory of an Exchange Economy”, *American Economic Review*, Volume 61, Number 5, pp. 784–805.
- Cannan, Edwin. (1921) “The Meaning of Bank Deposits”, *Economica*, Volume 1, Number 1, pp. 28–36.
- Cencini, Alvaro. (2001) *Monetary Macroeconomics, A New Approach*. London and New York: Routledge.
- Clower, Robert W. (1967) “A Reconsideration of the Microfoundations of Monetary Theory”, *Western Economic Journal*, Volume 6, Number 1, pp. 1–8.
- Committee on Payment and Settlement Systems. (2003) *The Role of Central Bank Money in Payment Systems*. Basle: Bank for International Settlements.

- Davidson, Paul. (1972) *Money and the Real World*. New York: John Wiley and Sons. Second Edition, 1978.
- De Grauwe, Paul. (2001) (Editor) *The Political Economy of Monetary Union*. Cheltenham and Northampton: Edward Elgar.
- Dow, Sheila C. and John Smithin. (1999) "The Structure of Financial Markets and the 'First Principles' of Monetary Economics", *Scottish Journal of Political Economy*, Volume 46, Number 1, pp. 72–90.
- Dowd, Kevin. (2000) "The Invisible Hand and the Evolution of the Monetary System", in John Smithin (Editor), *What is Money?* London and New York: Routledge, pp. 139–56.
- Einzig, Paul. (1966) *Primitive Money in Its Ethnological, Historical and Economic Aspects*. New York: Pergamon Press (first published 1949).
- Fisher, Irving. (1911) *The Purchasing Power of Money: Its Determination and Relation to Credit, Interest and Crises*. New York: Macmillan, 1931.
- Fleming, John M. (1971) "On Exchange Rate Unification", *Economic Journal*, Volume 81, Number 323, pp. 467–88.
- Foley, Duncan. (1989) "Money in Economic Activity", in John Eatwell, Murray Milgate and Peter Newman (eds) *The New Palgrave: Money*. London and Basingstoke: Macmillan, pp. 248–62.
- Friedman, Milton. (1974) "A Theoretical Framework for Monetary Analysis", in Robert J. Gordon (ed.) *Milton Friedman's Monetary Framework: A Debate with His Critics*. Chicago and London: University of Chicago Press, pp. 1–62.
- Friedman, Milton. (1987) "Quantity Theory of Money", in John Eatwell, Murray Milgate and Peter Newman (eds) *The New Palgrave: A Dictionary of Economics*. London and Basingstoke: Macmillan, vol. IV, pp. 3–20.
- Geanakoplos, John. (1987) "Overlapping Generations Model of General Equilibrium", in John Eatwell, Murray Milgate and Peter Newman (eds) *The New Palgrave: A Dictionary of Economics*. London and Basingstoke: Macmillan, Volume III, pp. 767–79.
- Goodhart, Charles A.E. (1989) *Money, Information and Uncertainty*. London and Basingstoke: Macmillan, 2<sup>nd</sup> edition (first published 1975).
- Goodhart, Charles A.E. (1998) "The Two Concepts of Money: Implications for the Analysis of Optimal Currency Areas", *European Journal of Political Economy*, Volume 14, Number 3, pp. 407–32.
- Gordon, Wendell. (1997) "Job Assurance – The Job Guarantee Revisited", *Journal of Economic Issues*, Volume 31, Number 3, pp. 826–34.
- Graziani, Augusto. (2003) *The Monetary Theory of Production*. Cambridge: Cambridge University Press.
- Grierson, Philip. (1977) *The Origins of Money*. London: University of London.
- Handa, Jagdish. (2000) *Monetary Economics*. London and New York: Routledge.
- Harvey, Philip. (1989) *Securing the Right to Employment: Social Welfare Policy and the Unemployed in the United States*. Princeton: Princeton University Press.
- Hayek, Friedrich A. von. (1978) *Denationalisation of Money: The Argument Refined: An Analysis of the Theory and Practice of Concurrent Currencies*. London: Institute of Economic Affairs, 3<sup>rd</sup> edition, 1990.
- Heinsohn, Gunnar and Otto Steiger. (1983) "Private Property, Debts and Interest or:

- The Origin of Money and the Rise and Fall of Monetary Economies”, *Studi Economici*, Volume 38, Number 21, pp. 3–56.
- Heinsohn, Gunnar and Otto Steiger. (1994) “A Private Property Theory of Debts, Interest and Money”, *Économies et Sociétés*, Volume 28, Number 1–2, pp. 9–24.
- Hicks, John R. (1967) *Critical Essays in Monetary Theory*. Oxford: Clarendon Press.
- Hicks, John R. (1975) “Crisis ’75...?”, *Institute of Economic Affairs Occasional Paper*. Special Issue, Number 43.
- Ingham, Geoffrey. (1996) “Money is a Social Relation”, *Review of Social Economy*, Volume 54, Number 4, pp. 507–29.
- Ingham, Geoffrey. (2000) “‘Babylonian Madness’: On the Historical and Sociological Origins of Money”, in John Smithin (Editor), *What is Money?* London and New York: Routledge, pp. 16–41.
- Innes, A. Mitchell. (1913) “What is Money?”, *Banking Law Journal*, May, pp. 377–408.
- Jones, Robert A. (1976) “The Origin and Development of Media of Exchange”, *Journal of Political Economy*, Volume 84, Number 4, pp. 757–75.
- Kahn, Charles M. and William Roberds. (2002) “The Economics of Payment Finality”, *Federal Reserve Bank of Atlanta Economic Review*, Volume 87, Number 2, pp. 1–12.
- Kenen, Peter B. (1969) “The Theory of Optimum Currency Areas: An Eclectic View”, in Robert A. Mundell and Alexander K. Swoboda (eds) *Monetary Problems of the International Economy*. Chicago and London: University of Chicago Press, pp. 41–60.
- Keynes, John M. (1930) *A Treatise on Money*, in *The Collected Writings of John Maynard Keynes*. London and Basingstoke: Macmillan, vols V–VI, 1971.
- Keynes, John M. (1936) *The General Theory of Employment, Interest and Money*, in *The Collected Writings of John Maynard Keynes*. London and Basingstoke: Macmillan, vol. VII, 1973.
- King, Robert G. and Charles I. Plosser. (1984) “Money, Credit, and Prices in a Real Business Cycle”, *American Economic Review*, Volume 74, Number 3, pp. 363–80.
- Kiyotaki, Nobuhiro and Randall Wright. (1989) “On Money as a Medium of Exchange”, *Journal of Political Economy*, Volume 97, Number 4, pp. 927–54.
- Kiyotaki, Nobuhiro and Randall Wright. (1991) “A Contribution to the Pure Theory of Money”, *Journal of Economic Theory*, Volume 53, Number 2, pp. 215–35.
- Kiyotaki, Nobuhiro and Randall Wright. (1993) “A Search-Theoretic Approach to Monetary Economics”, *American Economic Review*, Volume 83, Number 1, pp. 63–77.
- Klein, Benjamin. (1974) “The Competitive Supply of Money”, *Journal of Money, Credit, and Banking*, Volume 6, Number 4, pp. 423–53.
- Knapp, Georg F. (1924) *The State Theory of Money*, abridged edition translated by H.M. Lucas and J. Bonar. London: Macmillan (first German edition 1905).
- Kocherlakota, Narayana R. (1998) “Money is Memory”, *Journal of Economic Theory*, Volume 81, Number 2, pp. 232–51.
- Lerner, Abba P. (1943) “Functional Finance and the Federal Debt”, *Social Research*, Volume 10, Number 1, pp. 38–51.
- Lerner, Abba P. (1947) “Money as a Creature of the State”, *American*

- Economic Review*, Volume 37, Number 2, pp. 312–17.
- Malinowski, Bronislaw. (1921) “The Primitive Economics of Trobriand Islanders”, *Economic Journal*, Volume 31, Number 121, pp. 1–16.
- McCallum, Bennett T. (1987) “The Optimal Inflation Rate in an Overlapping-Generations Economy with Land”, in William A. Barnett and Kenneth J. Singleton (eds) *New Approaches to Monetary Economics*. Cambridge: Cambridge University Press, pp. 325–39.
- McCallum, Bennett T. (1990) “Inflation: Theory and Evidence”, in Benjamin M. Friedman and Frank H. Hahn (eds) *Handbook of Monetary Economics*. Amsterdam: Elsevier Science, vol. II, pp. 963–1012.
- McKinnon, Ronald I. (1963) “Optimum Currency Areas”, *American Economic Review*, Volume 53, Number 4, pp. 717–25.
- Menger, Karl. (1892) “On the Origin of Money”, *Economic Journal*, Volume 2, Number 6, pp. 239–55.
- Menger, Karl. (1923) *Grundsätze der Volkswirtschaftslehre*. Vienna: Hölder–Pichler–Tempsky (first published 1871).
- Minsky, Hyman P. (1986) *Stabilizing an Unstable Economy*. New Haven: Yale University Press.
- Mundell, Robert A. (1961) “A Theory of Optimum Currency Areas”, *American Economic Review*, Volume 51, Number 4, pp. 657–65.
- Niehans, John. (1978) *The Theory of Money*. Baltimore and London: Johns Hopkins University Press.
- Ostroy, Joseph M. (1973) “The Informational Efficiency of Monetary Exchange”, *American Economic Review*, Volume 63, Number 4, pp. 597–610.
- Pesek, Boris P. and Thomas R. Saving. (1968) *The Foundations of Money and Banking*. London and Basingstoke: Macmillan.
- Polanyi, Karl. (1944) *The Great Transformation: The Political and Economic Origins of Our Time*. New York: Farrar & Rinehart.
- Polanyi, Karl. (1977) *The Livelihood of Man*. New York: Academic Press.
- Realfonzo, Riccardo. (1998) *Money and Banking: Theory and Debate (1900–1940)*. Cheltenham and Northampton: Edward Elgar.
- Ricardo, David. (1817) *On the Principles of Political Economy and Taxation*. Cambridge: Cambridge University Press, 1951.
- Rochon, Louis-Philippe and Mario Seccareccia. (Editors), (2003) *Dollarization: Lessons from Europe and the Americas*. London and New York: Routledge.
- Rochon, Louis-Philippe and Matias Vernengo. (2003) “State Money and the Real World: or Chartalism and Its Discontents”, *Journal of Post Keynesian Economics*, Volume 26, Number 1, pp. 57–67.
- Schumpeter, Joseph A. (1954) *History of Economic Analysis*. London: Routledge, 1994.
- Selgin, George A. and Lawrence H. White. (1994) “How Would the Invisible Hand Handle Money?”, *Journal of Economic Literature*, Volume 32, Number 4, pp. 1718–49.
- Simmel, Georg. (1907) *The Philosophy of Money*. London: Routledge and Kegan Paul, 1978.
- Smith, Adam. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford: Clarendon Press, 1976.

- Spindt, Paul A. (1985) "Money Is What Money Does: Monetary Aggregation and the Equation of Exchange", *Journal of Political Economy*, Volume 93, Number 1, pp. 175–204.
- Tobin, James and Stephen S. Golub. (1998) *Money, Credit and Capital*. Boston: McGraw-Hill.
- Wallace, Neil. (1980) "The Overlapping Generations Model of Fiat Money", in John H. Kareken and Neil Wallace (eds) *Models of Monetary Economies*. Minneapolis: Federal Reserve Bank of Minneapolis, pp. 49–82.
- Walras, Léon. (1874) *Elements of Pure Economics*. London: Allen & Unwin, 1984.
- Woodford, Michael. (1990) "The Optimum Quantity of Money", in Benjamin M. Friedman and Frank H. Hahn (Editors), *Handbook of Monetary Economics*. Amsterdam: Elsevier Science, vol. II, pp. 1067–152.
- Wray, L. Randall. (1998) *Understanding Modern Money: The Key to Full Employment and Price Stability*. Cheltenham and Northampton: Edward Elgar.
- Wray, L. Randall. (2003) "The Neo-Chartalist Approach to Money", in Stephanie A. Bell and Edward J. Nell (Editors) *The State, the Market and the Euro: Chartalism versus Metallism in the Theory of Money*. Cheltenham and Northampton: Edward Elgar, pp. 89–110.
- Zazzaro, Alberto. (2003) "How Heterodox is the Heterodoxy of Monetary Circuit Theory? The Nature of Money and the Microeconomics of the Circuit", in Louis-Philippe Rochon and Sergio Rossi (eds) *Modern Theories of Money: The Nature and Role of Money in Capitalist*

*Economies*. Cheltenham and Northampton: Edward Elgar, pp. 219–45.

Sergio Rossi and Szymon Klimaszewski  
 Department of Political Economy  
 University of Fribourg  
 Switzerland  
 sergio.rossi@unifr.ch

## Cost Benefit Analysis

*Margaret Giles*

### **Introduction**

Cost benefit analysis (CBA) is a framework or technique for comparing the costs and benefits of a project or investment. Anthony Atkinson and Joseph Stiglitz (1980:475) argue that, “in principle, cost-benefit analysis is straightforward. Any investment project can be viewed as representing a perturbation of the economy from what it would have been had the project not been undertaken”.

CBA is the tool most often used by governments as a means of evaluating projects funded by taxpayers and/or that can affect communities of voters. It serves a similar purpose for businesses whose shareholders and/or management require ‘a financial analysis of a potential investment in order to determine its impact on the firm’s balance sheet and therefore on dividends and bonuses’ (Perkins 1994:3).

Most public finance textbooks contain a summary, if not a chapter, on CBA, usually with an applied example (see, for example, Auerbach & Feldstein 2002; Bailey 1995; Brent 1996; Brown & Jackson 1990; Browning & Browning 1987; Hare 1988; Hyman 1999; Jha 1998; Rosen 1999; Stiglitz 1988). Some texts (edited or synthesised) are specifically devoted to CBA (see, for example, Abelson 1986; Adler & Posner 2001; Bateman et al 2003; Button & Hensher 2005; Hanley & Spash 1993; Harberger & Jenkins 2002; Perkins, 1994; Watkins 2003).

### **Theoretical Background**

CBA has its theoretical origins in neoclassical welfare economics, namely,

Paretian welfare improvements (Pareto, 1894,1909) and, the extension to this, potential welfare improvements (the compensation principle) (Hicks 1939; Kaldor 1939). In welfare terms, a Pareto improvement occurs if, in the presence of a change, at least one person is made better off without any other person being made worse off. This restrictive rule was subsequently replaced with Hicks-Kaldor’s proposition that a potential improvement in welfare is sufficient to recommend the evaluated change or improvement. That is, if those who are likely to gain from the change or improvement can theoretically compensate those who might lose, and still be better off, then welfare is potentially improved.

This proposition leads directly to the technique of CBA. That is, projects or investments for which the benefits (the returns to those who gain from the change) outweigh the costs (the adverse effects on those who lose) are recommended; welfare can, theoretically, be improved. Changes for which the costs outweigh the benefits are not recommended; welfare is not likely to be improved. Importantly, this interpretation of Paretian welfare improvements allows for the gainers and losers to be different people.

The practical origins of CBA can probably be found at L’Ecole des Ponts et Chaussées (School of Civil Engineering) in Paris (established in 1747) which concerned itself with the economic evaluation of public works. Jules Dupuit, a graduate, turned his attention to the problem of measuring the social benefits of public works such as the provision of drinking water, and the building of roads, canals and bridges (Dupuit 1844). What Dupuit lacked in his approach though was a concise and practical methodology.

Frances Perkins (1994) suggests that such a methodology was not developed until the

1930s when the US Army Corps of Engineers was deliberating water supply alternatives across some western and central states. Whilst there was some discussion of methodology in earlier by Valdimir Key (1940) and Verne Lewis (1952), little rigour was given to CBA's methodology until, some forty years later, Ian Little & James Mirrlees (1969,1974) provided practical criteria for evaluating development projects.

Subsequently criticised by Amartya Sen and others for its emphasis on tradeable goods, the Little & Mirrlees methodology nonetheless underlies modern CBA. Sen (1970) exposed two issues. First, he argued that relative prices are not independent of the income distribution. Second, he suggested that value judgements are involved in the choice of shadow prices. In his approach, Sen therefore incorporates constraints, such as trade restrictions and other transaction costs, which may apply to the agency examining the new investment.

Typical projects that might be evaluated using CBA are highway, bridge or road building projects (see Abelson, 1986 for a useful Australian case study), and new schools, airports or defence equipment (including vessels, aircraft and military hardware) purchases. Consolidation or expansion of plant could also be appraised using CBA. The designation of a third London airport at Stansted, approved in 1985, was evaluated using CBA. Some of the controversy surrounding that project pertained to the application of the technique as much as to the project itself.

There is a standard approach to undertaking a CBA that comprehends the Little and Mirrlees criteria. This approach can be defined in terms of a series of consecutive steps. These steps are summarised below.

An example of a CBA of a government funded project (community patrol services in Western Australia (Giles 2003b) is used to illustrate each step. Whilst there is no doubt in the minds of community justice, health and education professionals, that community patrols are 'a good thing', governments that provide most, if not all, of their funds require a rigorous appraisal of their net worth to continue the funding or to expand their operations. This requirement is, as mentioned earlier, essential for the responsible use of taxpayer funds and good governance.

This project is different to the type of projects for which CBA is typically used thus highlighting some of the challenges and problems of CBA.

### **Defining Project Scope**

In the first step, the project, investment or change has to be outlined in terms of its scope, objectives and constraints. Scoping is about defining the impact of the project—answering the question “what is the difference in the state of the world with and without the project?” (Watkins 2003). This question can be addressed by determining the objectives of a project and identifying stakeholders. The objectives may depend on whether the project is small- or large-scale, and on whether constraints exist. The scoping will also provide a sense of the alternatives, the most common of which is to do nothing.

Generally, projects or investments are defined with a geographical constraint related to the jurisdiction of the organisation promoting the project. For example, a state or county government might want to consider the provision of a new public good within its sphere of influence. A firm with global resources servicing a global market

might be considering an investment beyond national borders. An aid agency might seek to provide tied grants to local communities on the basis that the costs and benefits are clearly identified locally. In this respect, choice of geographical boundary is arbitrary and CBA becomes a technique of partial equilibrium analysis.

Community patrols are defined as "non-coercive community intervention, or order maintenance, services designed to prevent or stop harm, and maintain community peace, security and safety" (Blagg & Valuri 2002). They have been introduced in indigenous and non-indigenous communities in Australia (from the late 1980s) and overseas. Whilst many existing patrols in Western Australia target indigenous issues in rural and remote communities (outside urban areas, with relatively small populations and geographically distant from major urban centres), there are also city-based patrols. For example, in the coastal capital city of Western Australia, Perth, the Noongar patrol is active. About one-third of patrols operating in Australia include non-indigenous youth and adults in their target groups.

In rural and remote communities in Western Australia, community patrol services have clearly defined operational/geographical boundaries. In urban areas, these boundaries are less clear but need to be delineated so that finite costs and benefits can be identified and valued (as discussed later).

The scope of a community patrol service is defined by the geographical area or community in which it will operate. The objectives for different communities may be generally similar but could also be idiosyncratic. Constraints are generally budgetary, but might also include

operational considerations peculiar to different communities.

### *Project Alternatives*

Importantly, the CBA should consider whether there are alternative means of achieving the project objectives and whether changing the size of the project might deliver outcomes that are as efficient and effective. Thus, it is not sufficient for a project to show net benefits. Its merits in a relative sense are also required. The difficulty for analysts at this stage in the CBA process is to recognise realistic but inferior alternatives to the proposed project. This is a vital component of the CBA.

The alternative to community patrol services with their preventative and conciliatory approach is not to have them, thereby allowing existing justice, health education and welfare agencies to continue their reactionary and often confrontational policies and practices.

### **Expenditures and Impacts**

The third step in the CBA process is to determine the expenditures and actions that will make up the project costs and the impacts that will determine the benefits.

#### *Costs*

The direct costs might cover capital, operating and maintenance expenditures. There are two types of capital expenditure that may apply to a project – fixed capital such as land, buildings, equipment, vehicles and infrastructure. Start-up costs such as workforce training or trial production runs could also be included as fixed capital. Assets remaining at the end of the project can have their ‘scrap value’ added to capital costs in the final year as a negative value



(Perkins 1994:38). Importantly, this scrap value is an economic or opportunity cost estimate not a final year depreciation amount that might be included in an accounting balance sheet.

The second type of capital is working capital which is used to finance the operating expenses of the enterprise. Its inclusion in the CBA needs to be attended carefully to avoid double counting of expenses. Working capital can include cash reserves and the value of stocks or raw materials and spare parts. The source of working capital may require some of the project parameters to be defined in a certain way. That is, the funding body may be an active, not passive, stakeholder. The extent that this impacts on costs will be peculiar to the project being evaluated. For example, a project requiring State finances may be subject to a constraint that State workers or businesses be engaged on the project.

Operating costs relate to the hire of labour, the use of raw materials and energy and associated costs such as insurance and marketing. For example, an immunisation programme in a third world village requires appropriate advertising of its availability and incentives to encourage participation. Otherwise the anticipated flow of benefits is not achieved. Taxes are included as operating costs if applicable. Maintenance expenditure, including ongoing staff training or equipment should be added. The funding body or other stakeholders may have *de minimis* limits on this type of expenditure.

The costs of operating community patrol services include fixed capital costs (such as the purchase of a minibus) and operating costs (such as fuel, registration and vehicle maintenance costs, plus bus driver wages). In some communities, volunteer drivers are used. However, whilst their wages may be

nil, in the CBA, the opportunity costs of their time (generally priced at the prevailing wage rate) is included. Cost differences between community patrol services tend to be catchment population size and peculiarities of the community rather than efficiency differences. Hence it is important not to assume that a CBA for a specific community patrol service will produce the same result as a CBA for other community patrol services. Analogously, welfare economists confirm that a CBA developed using WTP estimates from one population sample cannot necessarily be attributed to any other population sample.

Other operational costs arising in the delivery of community patrol services considered in a CBA are patroller/drive and police training, the provision of government oversight to ensure accountable funding and the provision of complementary activities such as school holiday camps.

Groenewegen (1990) points out that any project might also have indirect costs. These are “secondary costs or by-products of the project: an example ... is that the project may bid up local labour costs and thereby raise the local cost structure; the creation of environmental nuisances such as smoke or other forms of pollution and the destruction of wildlife would be classed as indirect intangible costs” (page 83).

The provision of community patrol services in some communities may result in increased rather than decreased resource usage in some respects. For example, the demand for referral services (such as welfare and housing) might increase as at risk individuals are located by patrols.

### *Benefits*

The benefits may be pecuniary or non-pecuniary. The items, as well as their

expected longevity, are required. For example, a new electrified rail system will require rolling stock that may be designed to last thirty years. On the one hand, the cost of the stock can be depreciated evenly over thirty years. On the other hand, benefits to commuters may increase over time as road congestion and petrol prices increase.

The benefits of community patrol services are various. Patrols intervene to prevent the occurrence of an event and to prevent the escalation of events that could lead to Police involvement. This is particularly applicable to 'hotspots' and at 'hot times'. Resulting outcomes include reduced imprisonments, detentions and remands in custody, arrests and charges for adults and juveniles. Savings to the criminal justice system are not inconsiderable.

Other impacts attributable to community patrol services are reduced child maltreatment and domestic violence, decreased property crime, reduced deaths in custody, reduced number and severity of drink-driving road crashes, reduced police patrols, reduced fear of crime and increased community peace, security and safety impacting positively on local businesses and residents. Premature deaths (particularly from drug and alcohol abuse), hospital and psychiatric admissions, ambulance call-outs, accident and emergency out-patient attendances, and prescription drug usage will reduce. Juvenile truancy will fall, school retention rates will increase and sporting and recreational participation will expand.

Community patrols, in discouraging anti-social behaviours, are able to facilitate the interaction of local people with support services that can provide training or access to employers, thus indirectly reducing unemployment and homelessness.

## **Project Parameters**

Choosing a timeframe and discount rate is the fourth step in the CBA methodology. In projects with capital costs, the life of the capital often dictates the timeframe. For community patrol services, the depreciation period or life expectancy for the major asset, the mini bus, between 7 and 10 years, could be used to define the timeframe for the CBA.

The costs and benefits of the project or investment need to be determined for each year that the project is in place, up to some limit. Small scale projects may have clearly defined years of operation and accrual of benefits. Larger projects (both in scope and/or financially) tend to have longer timeframes. The timeframe for the analysis is important for two reasons. In order to compare costs and benefits across time periods, a discount rate is chosen. The longer the timeframe and the higher the discount rate, the smaller the impacts of future costs and benefits. For example, a project with a life of thirty years evaluated with a high discount rate, say 14 percent, could show no benefits beyond the seventh year.

Discount rates, like geographical boundaries, are to some extent an arbitrary imposition. They are necessary because a CBA is examining costs and benefit streams into the future. With future dollars being worth less than current dollars to the individual and to society, it is important to know how the monetary amounts are deflated over time. The general assumption is that whatever the discount rate chosen, it is applied as a constant. This suggests that "people prefer income in period 1 over that in period 2, by the same amount by which they prefer income in period 2 over income in period 3. This assumption can be relaxed,

however, if there is evidence to the contrary” (Perkins 1994:66,fn 3).

Two alternative approaches to choosing a discount rate are to use a derived marginal rate of time preference (a social discount rate) or to use the opportunity cost of capital. Neither of these rates is singular. That is, individuals can have different time preferences and the opportunity cost may differ across different alternative uses. Neither are these rates the same. The social discount rate may deviate from the private rate for distributional reasons (Atkinson and Stiglitz 1980:480).

The marginal rate of time preference is the rate at which an individual values today’s dollar over tomorrow’s dollar. For example, a person preferring ten dollars today rather than eleven dollars in a year has a marginal rate of time preference of ten percent. The opportunity cost of capital is the market interest rate. Choosing between these two measures often depends on the project, in particular, how it is being funded. In addition, different rates may be employed for different projects “if the distributional impact differs across capital goods” (Atkinson and Stiglitz 1980:480). Irrespective of the measure, a discount rate that is too low may result in the CBA recommending an unworthy project. A discount rate that is too high could result in a CBA rejecting a worthy project.

Western Australian government agencies are required to use discount rates that vary depending on the risk of the project. In the case of community patrol services, a discount rate of 11.0% (low risk) would be appropriate.

### **Assigning Monetary Values**

The fifth step in developing a CBA is to assign values to the costs and benefits. These

values are most often calculated in the currency of the place of origin of the project or investment. However, costs and benefits for global projects might be valued in US dollars or Euros. Alternatively, the source of financing might dictate the currency used in the CBA. For example, projects considered by the World Bank tend to be valued in US dollars. The choice of currency is largely irrelevant to the conduct and outcome of the CBA (prevailing exchange rates can be used to amend the currency choice and purchasing power parity rules applied to these over future time periods). However, it is important that local prices are used if local resources are used.

The values assigned to costs and benefits, irrespective of currency, are usually derived from market rates (Evans 2000). Fuel and labour, for example, will be priced at the local pump price and wage rate, respectively. Costs would normally be easily calculated. However, there may be distortions in local labour, capital and foreign exchange markets necessitating corrections in prices before their inclusion in the CBA. Shadow prices can be used to make these corrections. “The problem of finding the correct shadow prices ... arises from the existence of market imperfections and failures” (Atkinson and Stiglitz 1980:475).

The estimated costs of a community patrol service will depend on the number of mini buses and patrol personnel and the hours of operation of the service. Mileage will also be a consideration. Police and patroller training will need to be costed in terms of wages and other (facilities and materials) expenditure. Estimates of the increased number and type of arrests and the average cost of arrests will be included. The provision of government oversight will

require assumptions about what resources may be required and the price of these. The cost of impacts on referral services may be estimated and included with other costs or, excluded from the CBA but discussed following calculation of the bottom line.

For some communities, the use of volunteer patrol drivers is preferred and there may be a tendency not to include the shadow wage for this resource. For comparative purposes, particularly when ranking the implementation of services across a number of communities, valuing some resources (such as labour) as 'costless' or 'free' for one community and 'costly' for another is not appropriate.

Benefits, however, may be more difficult to assign values to because the market may be non-existent, or public good or externality characteristics might be distorting the market price. For example, one of the non-pecuniary benefits that might apply to a project or investment is an improvement to quality of life or a saving of life. Consensus on values for these is yet to emerge (Giles, 2003a). In this case, shadow pricing may be adopted by using revealed preference (Blomquist 1979; Samuelson 1947) or contingent valuation (Ciriacy-Wantrup 1952; Davis 1963; Hanley & Spash 1993) techniques.

The value of benefits of community patrol services are calculated from the type of impact and the size of the impact. Reductions in detentions, remands in custody, arrests, custodial sentences and police patrols will impact of the costs of local police services and State justice services. Reduced violent crimes, child abuse, property damage and domestic/family violence will impact on the community, the criminal justice system and the insurance industry. Community patrol services might

also prevent deaths in custody and premature deaths from alcohol and drug abuse. The costs of post mortems and coronial enquiries related to these deaths are not insubstantial. Reducing such deaths will therefore produce a considerable saving. The loss of life can also be valued and included (Giles 2003a). Similarly for reduced road crashes and fatalities.

In terms of the health industry, reductions in hospital admissions and lengths of stay, ambulance call-outs, accident and emergency attendances and psychiatric admissions and lengths of stay will lower costs for government funded health services and reduce premiums for private health cover.

Reduced juvenile truancy and increased school retention rates can be valued in terms of social returns to education. Increased involvement in sport and recreation can also be valued in terms of social returns. Other benefits of community patrol services such as reduced unemployment, reduced homelessness and increased business trade can be valued in terms of contribution to gross State product, quality of life and sales/turnover, respectively.

Importantly, not all benefits will accrue to every community that operates a community patrol service. In addition, the impacts and value of impacts will be peculiar to each community. Hence CBAs need to be community-specific; there will be no generic cost/benefit result.

### **The Bottom Line**

Once values are assigned to costs and benefits, the next step is to calculate the bottom line. This can be done using net present value (NPV) or internal rate of return (IRR) calculations. Net present value sums the benefits over the timeframe of the

project and subtracts the costs over the same period. Importantly, the assigned values of costs and benefits are discounted by the chosen discount rate to ensure equivalence of values for future years with values for the current or first year. The NPV formula is as follows:

$$NPV = \sum_{t=0}^T \frac{B_t - C_t}{(1+r)^t}$$

where  $t$  denotes the particular year of the timeframe of the project and ranges from 0 in the first year to  $T$  in the final year;  $B_t$  is the sum of the benefits in time  $t$ ;  $C_t$  is the sum of the costs in time  $t$ ; and  $r$  is the chosen discount rate.

The IRR formula is given by:

$$\sum_{t=0}^T \frac{B_t - C_t}{(1+k)^t} = 0$$

where  $t$ ,  $T$ ,  $B_t$  and  $C_t$  are as defined for the NPV formula; and  $k$  is the discount rate that allows the net present value sum to zero.

### Decision Rule

In the next step, the bottom line, either  $NPV$  or  $k$ , is compared with a decision rule. Is  $NPV > 0$ ? That is, do the benefits of the project outweigh its costs? Alternatively, is  $k > r$ ? That is, is the return on the project greater than the return that could be received by the funds if invested elsewhere or greater than the minimum rate at which an individual values today's dollar over tomorrow's dollar, depending on which discount rate is being applied? Irrespective of formula used, the decision should be the same. If the net present value of the project is positive, then its rate of return will exceed the market (weighted average) cost of capital (interest rate) or the rate of time preference. Similarly, if the rate of return that equates the benefit and cost streams (thus reducing NPV to zero) exceeds the market interest rate, then the NPV will be greater than zero.

Community patrol services that operate efficiently and effectively tend to have positive NPVs and rates of return above the market rate. Thus decisions to offer these services to at risk-communities are supported by CBAs. In some cases, the CBA will provide clear support for a community patrol service - the value of benefits being much greater than the costs. In other cases, the support may be marginal and the CBA proves sensitive to assumptions made about the deployment of resources and the size of the impacts. Hence, reworking the CBA for a range of possible scenarios (sensitivity testing) is sensible.

### Sensitivity Tests

The final step in the CBA process is to adjust the project parameters to test whether the bottom line results are sensitive at the margin. The size of the project, the discount rate, the timeframe and the assumptions related to the costs and benefits (including assumed growth and inflation rates) can be changed and the earlier decision in favour or against the project or investment might be overturned.

### Problems of CBA

Despite its long history as an evaluation tool, CBA is confounded by number of problems. One of these is the distributional impact of proposed changes. In his summary of this issue, David Courard-Hauri (2004:192) refers to "the fact that the desires of high-income individuals are necessarily preferred over those of low-income individuals". However he adds that "this bias is present throughout any market economy, and traditional economist point to standard arguments about Pareto efficiency and the ability to address these inequities through progressive tax systems as reasons to leave

income weights out of CBA" (see also Adler & Posner 2001; Frank 2001). Governments and firms can, and do, use unweighted losses and gains to their advantage.

A second problem with CBA—measures of costs and benefits—has already been discussed. In short, the crux of CBA is monetising the costs and benefits. Some of these costs and benefits involve market resources and their estimates can be based on market prices. However, some of the costs and benefits do not involve market resources, hence there are no market prices to adopt. Alternative means of approximating their values, such as using shadow prices, are required.

Even using shadow pricing, obtained via hedonic pricing or contingent valuation, is not without bias. "If we have a cohort of individuals with similar abilities, backgrounds, and skills those for whom monetary consumption is relatively more important will, in a functioning market, fill the higher-paid positions and will therefore have the greatest effect on the 'value' of goods" (Courard-Hauri 2004:195).

Another problem with CBA is that it does not necessarily recommend an optimal size of the undertaking or investment. The operational environment, including funding and demographic issues, will influence the size of the project and it may not be obvious whether this size is optimal.

Many CBAs are conducted for projects being considered by single agencies. Sometimes the scope of the project ignores the repercussions (particularly on costs) on other agencies or, more broadly, elsewhere in the economy. In the case of community patrol services, funded by one agency, the increased demands on referral services provided by other agencies may be ignored in the CBA. The solution to this may lie with

inter-agency collaboration. In Western Australia, inter-agency collaboration is formerly supported by government.

### **The Future of CBA**

David Courard-Hauri (2004:192) summarises that the "increased reliance on CBA in recent years stems largely from a desire to increase reliance on objective, quantitative methods for policy comparison and optimization". This predilection is unlikely to go away anytime soon. Indeed, despite the aforementioned problems, some of which remain unresolved satisfactorily, CBA continues to be the appraisal technique of choice. He suggests increasing the use of willingness-to-accept (WTA) in lieu of willingness-to-pay (WTP) in contingent valuation studies as a way of minimising the aforementioned bias in shadow pricing. Broadening definitions of income to include time and self-satisfaction as well as money might also reduce this bias.

The degree of sophistication of current data collection, encoding and analysis practices and computing software has enabled CBA to be applied in more diverse ways and to include refined sensitivity analyses. This allows closer attention to, *inter alia*, exposing distributional impacts and aligning social and private discount rates. However, this also means that CBA is open to abuse (Servatius & Ferguson 1997).

In terms of government accountability, if the steps outlined here are adhered to and documented thoroughly, then decisions based on the bottom line will generally stand up to public scrutiny. However, political considerations may take precedence and the results of a CBA, if unpalatable to policymakers or business leaders, may be suppressed. That is, a project with a negative NPV might still be adopted (see

Groenewegen, 1990: 85). Justification is often given in terms of the intangibles (costs and benefits) - the difficulties of determining and valuing their impacts.

### **Alternatives to CBA**

Cost benefit analysis is not the only method of analysing potential projects or investments. Other evaluation techniques included cost-effectiveness analysis (Potas, Vining, & Wilson 1990:12) and data envelopment analysis (Carrington 1997). Both these techniques require finer data than a CBA, in terms of both quality and availability. For example, a cost effectiveness approach (CEA) would require the proposed project to be compared with an alternative project aimed at achieving the same objectives. CEA is often used when the outcomes or impacts can only be quantified in physical, but not monetary, terms (Groenewegen 1990).

### *Internet Sites*

Cost-Benefit Analysis.

<http://management.about.com/money/a/CostBenefit.htm>

Cost-Benefit Analysis.

[www2.sjsu.edu/faculty/watkins/cba.htm](http://www2.sjsu.edu/faculty/watkins/cba.htm)

### **Selected References**

- Abelson, P. (1986) *The Economic Evaluation of Roads in Australia*. Sydney: Australian Professional Publications.
- Adler, M.D. and E.A. Posner. (2001) (Editors) *Cost-Benefit Analysis: Legal, Economic, and Philosophical Perspectives*. Chicago: University of Chicago Press.
- Atkinson, A.B. and J.E. Stiglitz. (1980) *Lectures on Public Economics*. London: McGraw-Hill.
- Auerbach, A.J. and M.S. Feldstein. (2002) (Editors) *Handbook of Public Economics*. Amsterdam: North-Holland.
- Bailey, S.J. (1995) *Public Sector Economics: Theory, Policy and Practice*. London: Macmillan Press Ltd.
- Bateman, I. J.; A.A. Lovett and J.S. Brainard. (2003) *Applied Environmental Economics: a GIS Approach to Cost-Benefit Analysis*. Cambridge: Cambridge University Press.
- Blagg, H. and G. Valuri. (2002) *Profiling Night Patrol Services in Australia*. Perth: Crime Research Centre, The University of Western Australia.
- Blomquist, G. (1979) "The Value of Life Saving: Implications for Consumption Activity", *Journal of Political Economy*, Volume 87, Number 3, pp. 540-558.
- Brent, R.J. (1996) *Applied Cost-Benefit Analysis*. Aldershot and Brookfield, Vermont: Edward Elgar.
- Brown, C.V. and P.M. Jackson. (1990) *Public Sector Economics*. Fourth Edition. Oxford: Basil Blackwell.
- Browning, E.K. and J.M. Browning. (1987) *Public Finance and the Price System*. New York: Macmillan Publishing Company.
- Button, K.J. and D.A. Hensher. (2005) (Editors) *Handbook of Transport Strategy, Policy and Institutions*. Amsterdam: Elsevier.
- Carrington, R. (1997) "Performance Measurement in Government Service Provision: The Case of Police Services in New South Wales", *Journal of Productivity Analysis*, Volume 8, Number 4, pp. 415 - 430.
- Ciriacy-Wantrup, S.V. (1952) *Resource Conservation: Economics and Policies*. Berkeley: University of California Press.
- Courard-Hauri, D. (2004) "The Effect of Income Choice on Bias in Policy

- Decisions Made Using Cost-Benefit Analyses”, *Ecological Economics*, Volume 51, pp. 191-199.
- Davis, R. (1963) “Recreation Planning as an Economic Problem”, *Natural Resources Journal*, Volume 3, Number 2, pp. 239 - 249.
- Dupuit, J. (1844) “On the Measurement of the Utility of Public Works”, *International Economic Papers*, Volume 2, 1952.
- Evans, A.W. (2000) “Economic Evaluation of Road Traffic Safety Measures: United Kingdom”, in U.E.R. Centre (Editor), *European Conference of Ministers of Transport*, Paris, pp. 77-99.
- Frank, R.H. (2001) “Why is Cost-Benefit Analysis so Controversial?”, in M.D. Adler and E.A. Posner (Editors), *Cost-Benefit Analysis: Legal, Economic, and Philosophical Perspectives*. Chicago: University of Chicago Press.
- Giles, M. (2003a) “The Cost of Road Crash Costs: A Comparison of Methods and Recent Australian Estimates”, *Journal of Transport Economics and Policy*, Volume 37, Number 1, pp. 95-110.
- Giles, M. (2003b) *A Framework for a Cost-Benefit Analysis of Community Patrol Services*. Perth: Report to the Crime Research Centre, The University of Western Australia.
- Groenewegen, P. (1990) *Public Finance in Australia: Theory and Practice*. Third Edition. Sydney: Prentice Hall.
- Hanley, N. and C.L. Spash. (1993) *Cost-Benefit Analysis and the Environment*. Hants, England: Edward Elgar.
- Harberger, A.C. and G.P. Jenkins. (2002) *Cost-Benefit Analysis*. Cheltenham, UK: Edward Elgar.
- Hare, P.G. (1988) (Editor) *Surveys in Public Sector Economics*. Oxford: Blackwells.
- Hicks, J. (1939) Foundations of Welfare Economics. *Economic Journal*, Volume 49, Number 196, pp. 696-712.
- Hyman, D.N. (1999) *Public Finance: a Contemporary Application of Theory to Policy*. Sixth Edition. Fort Worth, Texas: The Dryden Press.
- Jha, R. (1998) *Modern Public Economics*. London: Routledge.
- Kaldor, N. (1939) “Welfare Propositions in Economics and Inter-Personal Comparisons of Utility”, *Economic Journal*, Volume 49, Number 195, pp. 522-549.
- Key, V.O. (1940) “The Lack of Budgetary Theory”, *American Political Science Review*, Volume 34, Number 6, pp. 1137-1140.
- Lewis, V.B. (1952) “Toward a Theory of Budgeting”, *Public Administration Review*, Volume 12, Number 1, pp. 43-54.
- Little, I.M.D. and J.A. Mirrlees. (1969) *Manual for Industrial Analysis in Developing Countries, II, Social Cost Benefit Analysis*. Paris: OECD.
- Little, I.M.D. and J.A. Mirrlees. (1974) *Project Appraisal and Planning for Developing Countries*. London: Heinemann Educational Books.
- Pareto, V. (1894) “Il Massimo di Utilita dato dalla Libera Concorrenza”, in G. Busino (Editor), *Ouvres Completes Tome XXVI: Ecrites D'Economie Politique*. Geneve: Libraire Droz, 1982, pp. 276-294.
- Pareto, V. (1909) *Manual of Political Economy [Manuel d'Economie Politique]*. New York: Augustus M Kelly Publishers, 1971.
- Perkins, F. (1994) *Practical Cost Benefit Analysis: Basic Concepts and Applications*. South Melbourne: MacMillan Education Australia Pty Ltd.



- Potas, I., A. Vining and P. Wilson. (1990) *Young People and Crime: Costs and Prevention*. (Report to the Department of Employment, Education and Training (Youth Bureau) Canberra: Australian Institute of Criminology.
- Rosen, H.S. (1999) *Public Finance*. Fifth Edition. Boston Burr Ridge, Illinois: Irwin McGraw-Hill.
- Samuelson, P.A. (1947) *Foundations of Economic Analysis*. Cambridge, Mass.: Harvard University Press.
- Sen, A.K. (1970) *Collective Choice and Social Welfare*. London: Oliver and Boyd.
- Servatius, B. and C.W. Ferguson. (1997) *AM/FM/GIS Investment Analysis*. March 23-36. Paper presented at the Geospatial Information and Technology Association 20th Annual Conference, Nashville, Tennessee.
- Stiglitz, J E. (1988) *Economics of the Public Sector*. Second Edition. New York: W.W. Norton.
- Watkins, T. (2003) *Introduction to Cost Benefit Analysis*.  
[www2.sjsu.edu/faculty/watkins/cba.htm](http://www2.sjsu.edu/faculty/watkins/cba.htm)

*Margaret Giles  
School of Business  
Edith Cowan University  
Perth, Australia  
m.giles@ecu.edu.au*

## Debt Deflation

Reynold Nesiba

### Introduction

Debt-deflation refers to the vicious downward spiral created when an over-indebted and layered financial system faces a period of falling prices. Firms and individuals confronted with excessive debt and falling prices and incomes find themselves unable to service existing debt commitments. They react by defaulting or selling marketable assets. These large and sudden attempts to reduce over-indebtedness cause further deflation not only in consumer prices but also in asset prices (stocks, bonds, real estate, etc.) leading to widespread default and bankruptcy. One paradoxical feature of this process is that the collective attempts by debtors to reduce debt instead cause price levels to fall faster than nominal debt levels, and thereby increase real (inflation-adjusted or commodity-based) debt burdens. This cumulative process is thought to have played a central role in the cause of the Great Depression of the 1930s, particularly in the US, as well as the more recent decade of decline in Japan. The reemergence of deflation or low inflation in Japan, Germany, and the US in the early 2000s has led to renewed interest in this 70-year old idea.

The term debt-deflation viewed singly has led some students to misleadingly believe it refers to a period where debt *decreases* in value. This is incorrect. One intermediate result of debt-deflation is that the real or commodity value of debt and debt servicing payments *increase* relative to the value of other assets, incomes, and commodity values—despite decreasing in nominal terms. A more accurate view of

debt-deflation considers its reference to two separate but interrelated economic conditions of 1) high indebtedness, and 2) falling overall price levels. Each of these conditions alone can be problematic for a capitalist economy. However, when combined, high indebtedness and a sustained deflation can become in Irving Fisher's (1933:341) words, "the big bad actors" in creating economic panics and crashes.

### Hypothetical Example

A simplified and personal example beginning with a single borrower may make this concept easier to grasp (adapted from Burton and Lombra, 2003:308). Assume that you have taken out a \$200,000 mortgage on a \$200,000 house at a fixed interest rate at 4.388% for 30 years. Your monthly payment is \$1,000 per month (\$12,000 per year) and your annual income is \$36,000. Thus, debt payments constitute one-third of your income. If real estate values and your income continue to increase, or simply remain steady, you will easily meet your payment commitments. Now consider how price deflation will affect your ability to manage this situation of high indebtedness.

For the sake of illustration, assume deflation causes your asset prices (home value), income, and other output prices to all fall by 50%. The value of the house is now worth \$100,000 and your income has fallen to \$18,000. In inflation-adjusted terms, you *appear* to be as well off as you were before the decrease in the overall price level. Despite your income falling by one-half, the prices of food, clothing, and other goods and services have on average also fallen in half. However, your mortgage contract, made in *nominal* terms, has not been reduced by this turn of events. Before the deflation, you could easily meet your payment

commitments when they required one-third of your income. Now that your income is reduced to \$18,000, your annual mortgage obligations of \$12,000 constitute two-thirds of your income. Given your other essential expenses for food, clothing, transportation, medical care, and so on, you can no longer meet your debt obligations. Deflation has caused your *real* debt burden to double as a share of your income. (For instance, you would now be required to work twice as many hours just to make your mortgage payment.) However, the *nominal* debt burden of \$1,000 per month remains unchanged. Clearly your debt obligations ensure that you are not as well off as you were before. You are in default because you cannot meet your debt requirements, and assuming no other assets or debts; you are insolvent because your liabilities (your mortgage) are greater than the value of your assets (your house). Falling nominal interest rates may appear to provide some relief for borrowers with variable interest rates. However, if *nominal* interest rates fall 25% in this example while other prices are falling 50%, the *real* interest rate actually increases at this worst of all possible times. Even if the interest rate fell in half to 2.194%, your mortgage payment would still be \$758.79 per month or \$9,105.48 per year—more than 50% of your new lower income. The banker who was so happy to originate your mortgage will now begin foreclosure proceedings. Your home will be repossessed, and you will become bankrupt.

If the banker who now owns your home attempts to sell it and recoup some of her loss, she will be unable to raise anywhere near the amount of your outstanding mortgage. Other borrowers and lenders with repossessed properties are experiencing similar problems. In a futile attempt to meet

their own debt commitments, they too are trying to sell their properties and/or previously originated mortgages. However, with a glutted real estate market (and market for mortgages) and few if any buyers, prices decline further, incomes fall further, output and employment decline, and the downward spiral continues with more defaults, foreclosures, and bankruptcies. At financial institutions, the losses from bad loans continue to increase. If the banking system lacks deposit insurance, depositors may attempt to withdraw their funds leading to widespread bank runs. Without a central bank to serve as a lender of last resort and/or a large central government to reinvigorate aggregate demand with deficit spending, the economy may slide into panic and crisis.

Although this example is purposely simplified and involves an unrealistically large decrease in prices, one can see the self-reinforcing problems caused by deflation in the context of high indebtedness as well as the need for prompt and effective expansionary fiscal and monetary policy.

### **Intellectual History**

According to Wolfson (1994:13-16), theorists who have written about various kinds of economic and financial crises go back at least to Karl Marx whose first volume of *Capital* was published in 1867. Marx's "falling rate of profit" which is realized as a "money crisis," can be seen as a precursor to debt-deflation theory. By 1904, Chapter VII of Thorstein Veblen's *The Theory of Business Enterprise* (p. 192) describes a process where "a period of liquidation" results in the cascading fall of loans, decreases in the value of underlying collateral, leading to liquidations. Veblen's student, Wesley Clair Mitchell (1913), went on to develop a sophisticated theory of

business cycle phenomena including a financial crisis theory characterized by rising debt levels over the course of the business cycle, rising input costs, higher long-term interest rates, and an increased reliance on short-term debt. Declining profits and higher interest rates eventually undermine business profits leading to debt liquidation and crisis.

Despite Marx, Veblen, and Mitchell's prior work on related issues, the best-known proponent of debt-deflation theory and the originator of this hyphenated phrase is Irving Fisher. He developed this theory while teaching at Yale University in 1931 and in a book titled *Booms and Depressions* in 1932. The most famous exposition of this idea was published a year later in *Econometrica* as the "Debt-Deflation Theory of Great Depressions." Here Fisher explains a nine-step interactive process:

"Assuming, accordingly, that, at some point of time, a state of over-indebtedness exists, this will tend to lead to liquidation, through the alarm either of debtors or creditors or both. Then we may deduce the following chain of consequences in nine links: (1) *Debt liquidation* leads to *distress selling* and to (2) *Contraction of deposit currency*, as bank loans are paid off, and to a slowing down of velocity of circulation. This contraction of deposits and of their velocity, precipitated by distress selling, causes (3) *a fall in the level of prices*, in other words a swelling of the dollar. Assuming, as above stated, that this fall of prices is not interfered with by reflation or otherwise, there must be (4) *A still greater fall in the net worths of businesses*, precipitating bankruptcies and (5) *A like fall in profits*, which in a "capitalistic," that is, a private-profit

society, leads the concerns which are running at a loss to make (6) *A reduction in output, in trade and in employment* of labor. These losses, bankruptcies, and unemployment, lead to (7) *Pessimism and loss of confidence*, which in turn lead to (8) *Hoarding and slowing down still more the velocity of circulation*. The above eight changes cause (9) *Complicated disturbances in the rates of interest*, in particular, a fall in the nominal, or money, rates and a rise in the real, or commodity, rates of interest." [Fisher 1933-34:341-342]

Fisher explains the ironic results of this process two pages later (emphasis in the original):

"Each dollar of debt still unpaid becomes a bigger dollar, and if the over-indebtedness with which we started was great enough, the liquidation of debts cannot keep up with the fall of prices, which it causes. In that case, the liquidation defeats itself. While it diminishes the number of dollars owed, it may not do so as fast as it increases the value of each dollar owed. Then, *the very effort of individuals to less their burden of debts increases it, because of the mass effect of the stampede to liquidate in swelling each dollar owed*. Then we have the great paradox which, I submit, is the chief secret of most, if not all, great depressions: *The more the debtors pay, the more they owe*. The more the economic boat tips, the more it tends to tip. It is not tending to right itself, but is capsizing." [Fisher 1933:344]

Over time, Fisher's insight that the twin problems of debt and falling prices are an inherent source of financial instability has

been extended and amended by his contemporary John Maynard Keynes (1936), and later theorists such as Hyman Minsky (1982, 1986, 1995), Martin Wolfson (1994, 1996, 2002), and Steve Keen (1998). Keynes' great work directly addresses the problems of a modern economy with a well-developed financial system that influences production in the non-financial sector. Hyman Minsky's Financial Instability Hypothesis builds on the work of both Fisher and Keynes, to explain how the normal functioning of an economy with sophisticated financial relations will endogenously generate business cycles and economic fluctuations. Minsky's key contribution is showing that over a period free of financial panics and deep recessions, capitalist financial systems change such that debt levels increase, borrowers become characterized by higher levels of risk, and the system as a whole becomes more vulnerable to a debt-deflation process. This matches the post-war US experience and provides an explanation for the "state of over-indebtedness" that Fisher simply assumes as a starting point.

Martin Wolfson's (1996:317-319) work explores the possibility "by which something akin to a debt-deflation process" could occur in the United States given the economic conditions of the late 1990s. Specifically he shows how Fisher's approach has been modified by others to consider that 1) the problems caused by deflation (which has not occurred for decades) are similar to those caused by decreases in the rate of inflation (which has characterized much of the period after the early 1980s). Wolfson also more carefully explores 2) the role played by the banks and the banking system, as well as 3) the problems of meeting payment commitments

that may be caused by a sudden decrease in financial asset prices. Finally, Steve Keen (1998) uses the mathematics of nonlinear dynamic modeling to illustrate how Minsky's theory can be examined in the context of a complex systems point of view.

The debt-deflation concept can also be applied to aid in understanding the international financial system. The Asian financial crisis of 1997-1998 illustrated that a debt-deflation process can be initiated by falling currency values when debt obligations are contracted in foreign currency terms. In the case of Indonesia for instance, the falling value of the rupiah dramatically increased its dollar-denominated debt obligations. The inability to meet these payments led to capital flight, falling aggregate expenditures, declining asset values, and thus further declines in the rupiah's value (Wolfson 2002).

### **Legislative & Governance Consequences**

Fisher and subsequent theorists have shown that a devastating debt-deflation process can be initiated by both domestic and international phenomena. Further, it can be caused not only by falling overall price levels, but also falling asset prices (such as a stock market crash) and/or falling currency values which, depending on economic conditions, may lead subsequently to substantial declines in real incomes and/or the creation of some economic units that are unable to meet their payment commitments. These disruptions in turn create an intense demand for liquidity by a variety of economic units (e.g., 1) investors who invested on margin, 2) risk arbitragers, 3) sellers of stock-indexed put options, 4) stock market mutual funds facing redemptions, 5) market specialists, and 6) brokerage houses.) Distressed selling of assets by these units

will further exacerbate the process Wolfson (1996:324-325). To prevent this downward spiral, policy makers must be vigilant and undertake three critical macroeconomic policy actions when this situation arises.

First, the monetary authorities must engage in decisive and aggressive monetary policy. This can be done through the central bank's purchase of securities that drives down interest rates and injects critical liquidity into a system desperate for cash.

Second the central bank must also stand ready to use its discount window to make emergency loans to the banking industry and to otherwise serve as a lender of last resort to the financial sector and perhaps even the non-financial sector.

Third, the legislative branch of the federal government needs to engage in discretionary fiscal policy to maintain aggregate demand and facilitate economic growth. This can be done by aggressively increasing government spending and/or by engaging in tax cuts.

In addition to the aggressive discretionary use of expansionary fiscal and monetary policy, we are fortunate that post-war western economies have institutionalized automatic stabilizers as part of big government capitalism. During a business cycle downturn, incomes fall and reduce government tax revenue. At the same time, the federal government increases its spending on various social welfare programs (such as unemployment insurance, cash assistance, housing subsidies, and food stamps). Thus, with no discretionary change in fiscal policy, the government budget tends toward deficit. This helps to cushion the decrease in aggregate spending and to maintain cash flows and profits above what they would otherwise be. During the business cycle upturn, automatic stabilizers

dampen the expansion. Incomes grow as do tax revenues. The federal government spends less on social welfare programs and the budget tends toward a surplus. This contractionary affect helps prevent the business cycle from going into a boom and dampens (but does not eliminate) the incentive for private firms to over-invest in productive capacity (Keen 1998:2).

### **Global Threats to Stability**

Although not an exhaustive list, there are three prominent threats to macroeconomic stability from a debt-deflation perspective. The first and foremost threat is the worship of balanced budgets. The description above highlights the importance that automatic stabilizers play during an economic downturn. If policy makers were forced to balance budgets during this downturn, they would be raising taxes and decreasing spending at the worst possible time and exacerbating the downward spiral. In 1995, the US Senate came within one vote of the necessary two-thirds needed to approve a balanced budget amendment to the constitution. In Europe, the treaty mechanism used to create and maintain the European Monetary Unit (EMU) requires strict guidelines with respect to budget deficits and has given a deflationary bias to the region. This is one of the reasons that Germany's rate of inflation fell below 1% and threatened to turn negative in 2003.

Second and related to the first, the Asian Crisis of the late 1990s made clear, that the one-size-fits-all austerity packages waged by the International Monetary Fund (IMF) are not effective. They may in fact be counterproductive in stemming deflation and further currency devaluation given the current institutional framework of the global financial system. Future "assistance" will

need to be tailored to the specific conditions and institutions of particular countries and likely need to err on the side of generosity rather than austerity.

Third and finally, the global economy lacks an institution that can effectively serve as a global lender of last resort (lender of last resort: International). Until such an institution is created, there is a need for greater macroeconomic coordination among the world's leading economic powers, especially in times of global financial crises.

After 70 years of relative obscurity, Fisher's debt-deflation notion and its extensions continue to illuminate public policy problems of international interest and to point policymakers toward productive solutions.

### Internet Sites

Japan Inc. *Japanese Business News*.  
[www.japaninc.net/newsletters/?list=mw&issue=8](http://www.japaninc.net/newsletters/?list=mw&issue=8)

### Selected References

- Burton, Maureen and Ray Lombra. (2003) *The Financial System and the Economy: Principles of Money and Banking*. Third Edition. South-Western College Publishing.
- Davidson, Paul. (1972) *Money and the Real World*. New York: Wiley.
- Dimand, Robert W. (1994) "Irving Fisher's Debt-Deflation Theory of Great Depressions" *Review of Social Economy*, Volume 52, pp. 92-107.
- Fisher, Irving. (1932) *Booms and Depressions*. New York: Adelphi Company.
- Fisher, Irving. (1933) "The Debt-Deflation Theory of Great Depressions," *Econometrica*, Volume 1, Issue 4, pp. 337-357.
- Keynes, John Maynard. (1936) *The General Theory of Employment, Interest, and Money*. New York: Harcourt, Brace & World Inc.
- Keen, Steve. (1998) "The Nonlinear Dynamics of Debt Deflation", *Complexity International*, Volume 6, pp. 78-93.
- Marx, Karl. (1894) *Capital*, Volume III. New York: International Publishers, 1967.
- Minsky, Hyman P. (1995) "Longer Waves in Financial Relations: Financial Factors in the more Severe Depressions II" *Journal of Economic Issues*, Volume 29, Issue 1, pp. 83-97.
- Minsky, Hyman P. (1986) *Stabilizing an Unstable Economy*. New Haven: Yale University Press.
- Minsky, Hyman P. (1982) *Can 'It' Happen Again? Essays on Instability and Finance*. Armonk, NY: M.E. Sharpe.
- Mitchell, Wesley Clair. (1913) *Business Cycles*. Berkley: University of California Press.
- Posen, Adam S. (2003) "Frog in the Pot: Germany's Path to the Japan Syndrome", *The National Interest*, Spring, pp. 105-117.
- Tobin, James. (1978) "A Proposal for International Monetary Reform", *Eastern Economic Journal*, pp. 153-159.
- Veblen, Thorstein. (1904) *Theory of Business Enterprise*. New York: Charles Scribner's Sons.
- Visano, Brenda Spotton. (2002) "Financial Manias and Panics: A Socioeconomic Perspective", *American Journal of Economics and Sociology*, Volume 61, pp. 801-826.
- Wolfson, Martin H. (2002) "Minsky's Theory of Financial Crises in a Global Context," *Journal of Economic Issues*, Volume 36, Number 2, pp. 393-400.

Wolfson, Martin H. (1996) "Irving Fisher's Debt-Deflation Theory: Its Relevance to Current Conditions", *Cambridge Journal of Economics*, Volume 20, pp. 315-333.

Wolfson, Martin H. (1994) *Financial Crises: Understanding the Postwar U.S. Experience*. Second Edition. Armonk, NY: M.E. Sharpe.

*Reynold F. Nesiba*  
*Department of Economics*  
*Augustana College*  
*South Dakota, USA*  
*reynold.nesiba@augie.edu*



## Deposit Insurance

*Kam Hon Chu*

### **Introduction**

Incorporating explicit deposit insurance schemes is a recent global trend, spreading from industrialized countries to emerging economies. Deposit insurance is not novel. Its antecedent can at least be traced back to the New York Safety Fund in 1829, although the first nation-wide, publicly-run scheme -- the Federal Deposit Insurance Corporation -- was established in 1934 as an aftermath of the Great Depression. The number of such schemes worldwide did not grow rapidly until the mid-1980s when many countries began to be plagued by waves of economic and financial crises. By 2000 a total of 76 countries have adopted these schemes in one form or another (Demirgüç-Kunt and Kane 2002), while some countries are developing or seriously considering introducing similar schemes.

Deposit insurance is a guarantee offered by an insurer, usually the government, to depositors to cover their deposits up to a specified amount when their deposit-taking institutions (mainly banks) go bankrupt. Deposit insurance lowers the cost of funds to depository institutions because depositors perceive their insured deposits as virtually risk-free and are therefore satisfied receiving interest close to the risk-free rate. The insurance premium can be mispriced if it is not actuarially fair to reflect insured institutions' risk exposure, as in the case of a flat premium. Like other forms of insurance, deposit insurance affects incentives of both depositors and depository institutions because insurance risk and assessment change the trade-off between risk and expected returns. Therefore it has far-

reaching and profound impacts on corporate governance as well as monetary and financial governance.

One of banks' major functions is to provide monitoring and information-related services. According to the principal-agent theory, information asymmetry causes contracting problems between lenders and borrowers. The delegation of screening and monitoring borrowers to banks is an efficient allocation mechanism that substantially reduces screening and monitoring costs (Diamond 1984). Simply put, banks play a crucial role in corporate governance. This is notably the case in countries like Japan (main bank system) and Germany (hausbank system) where banks, holding both debts and equities of their client firms, play a corporate governance role by acting as outside monitors to alleviate the agency problem so as to ensure that managers pursue shareholders' interests.

Of course, this raises the problem "who monitors the monitor?" as banks themselves also face a similar agency problem among their managers, shareholders and depositors. Indeed, most, if not all, issues associated with deposit insurance intertwine with the debates on financial regulation and deregulation and on central banking versus free banking.

### **The Case for Deposit Insurance**

Based on the public-interest theory, protection of small depositors is a principal and "official" argument for deposit insurance. Small depositors usually do not have financial knowledge or skills to monitor banks; or it is too costly for them to do so individually, i.e., the total monitoring costs outweigh the benefits. By delegating the responsibility of monitoring banks to the deposit insurer, the monitoring costs would

be substantially reduced to attain net social savings.

However, systemic stability rather than protection of small depositors is widely accepted as the primary goal of deposit insurance (Golembe 1960), because deposit insurance prevents substantial welfare losses due to bank runs (e.g. Diamond & Dybvig 1983). Without deposit insurance, depositors, particularly small depositors, may incline to panic in response to bad news, like a bank failure, or even rumours. Given information asymmetry and banks' first-come, first-served basis of paying out cash reserves, depositors' rational response is to withdraw their deposits when they lose confidence in their banks. Their simultaneous actions trigger bank runs and contagions occur if financially sound banks are also afflicted. Bank runs are presupposedly bad because they interrupt financial intermediation by forcing premature liquidations of productive investment projects. Theoretically bank runs may drive economically sound banks into insolvency because of fire-sale losses, although they are rare in reality (Kaufman 2000). A fractional banking system that channels short-term deposits into long-term investment is presumed to be structurally vulnerable to contagious runs. This contagion or domino effect is a prime example of third-party effects or externalities to justify government intervention. By guaranteeing the safety of deposits, deposit insurance promotes banking stability by eliminating runs and reducing systemic risk.

Like bank runs, bank failures have deleterious effects on the financial system and the real sector because of banks' pivotal role in financial intermediation. Bank failures are costly because information is

lost on borrowers, who have to pay more expensive credit terms when they turn to solvent banks. More importantly, the insolvency of groups of banks could lead to macro-domino externalities such as low investment and high unemployment, as evidenced in the Great Depression (Bernake 1983).

Uncertain about the political and economic consequences of bank failures, especially in cases involving large banks, regulators in countries without explicit deposit insurance schemes incline to follow the too-big-to-fail doctrine and act as a lender of last resort when banks are in financial distress (Bordo 1992). Under these circumstances, regulators essentially offer unlimited implicit deposit insurance to banks. By contrast, an explicit deposit insurance scheme limits the extent of coverage provided by the government in case of bank failures. Furthermore, the government's burden is relieved by banks' contributions to the insurance fund for resolutions of their failed peers.

A closely related, perhaps more political than economic, reason for instituting an explicit deposit insurance scheme is that the government can avoid political battling with small depositors when a bank failure actually occurs (Kaufman 1996).

From small banks' perspectives, the too-big-to-fail doctrine gives large banks an unfair competitive advantage because regulators are virtually providing implicit deposit insurance to large banks only. As such, depositors tend to keep their deposits at large banks, driving the industry towards an oligopoly and hence deviations from allocation efficiency. To promote competition and efficiency, explicit deposit insurance should thus be offered across the

board to level the playing field between small and large banks (Garcia 2000).

Banking stability and efficiency together lead to another argument for deposit insurance. Banking stability is regarded as a prerequisite of financial stability, which encourages savings and economic growth. Recent empirical evidence confirms banks' positive role in promoting economic growth (e.g. Levine et al Beck 2000). Deposit insurance has favourable impacts on financial stability and development in the presence of a strong institutional environment—i.e. the rule of law, accounting and auditing standards, quality of government, low corruption, etc.; otherwise it harms financial development (Cull et al 2002). Against a trend of globalization, a stable and efficient banking system is of vital importance in attracting and channelling both foreign capital and domestic savings into productive investment.

Finally, deposit insurance serves to help local banks in emerging economies to acquire and retain their market shares in deposits that might otherwise be taken up by large foreign-owned banks because of financial liberalization and globalization (World Bank 2001). To maintain international competitiveness and to prevent contagions from abroad are among the factors contributing to a worldwide trend in instituting explicit deposit insurance schemes.

### **The Case Against Deposit Insurance**

A principal argument against deposit insurance is that it incurs the notorious problems of moral hazard and adverse selection. Deposit insurance induces the moral-hazard problem, as depositors no longer scrutinize banks' financial conditions,

and creates an incentive for banks towards excessive risk-taking. Under limited liability laws, banks keep their profits if their speculation turns out to be profitable and the insurer absorbs the losses if the speculation fails. This perverse incentive is even stronger when a bank is at the brink of bankruptcy, because it gambles for resurrection by offering high deposit rates in order to attract deposits to finance additional loans. The higher deposit costs have to be covered by higher lending rates. This leads to the adverse-selection problem because borrowers who are willing to pay higher interest rates are more likely to have bad credit risks, thus increasing the bank's probability to fail. All these problems have long been recognized by economists and policymakers, although sophisticated theoretical models demonstrate them formally are developed much later (e.g., Merton 1977).

In the US savings and loan and banking industries, the large numbers of failures and substantial losses to their insurance funds during the 1980s (see e.g., Kane 1985, 1989 for details) convincingly demonstrate that deposit insurance weakens market discipline and aggravates systemic risk if the moral-hazard and adverse-selection problems are not contained. To achieve the latter goal stringent regulations such as capital adequacy requirements are needed, at the expense of regulatory costs and distortions to efficient allocation of resources. By contrast, there are less costly devices for maintaining banking stability and protecting depositors, such as narrow banks (Litan 1987), chequable money market mutual funds (Cowen & Kroszner 1990), government saving bonds (Chu 2000), to name just a few.

Besides its huge costs, the effectiveness of deposit insurance in eliminating bank runs in practice is challenged. If the insurance fund is, or is perceived to be, inadequate to cover potential losses from bank failures, depositors still have incentives to run on banks, as witnessed in Ohio and Maryland in 1985 (Kane 1989). Furthermore, large banks that serve largely uninsured depositors can still be subject to runs or “walks,” as evidenced in the collapse of Continental Illinois in 1984 (Sprague 1986).

Even if deposit insurance actually reduces the likelihood of bank runs, the argument for deposit insurance has highlighted the bad effects of runs but ignored their good effect – the strong market discipline on bank management. Demandable deposit contracts provide incentives for banks to invest intelligently and prudently to avoid insolvency or misperceptions of being insolvent (Calomiris and Khan 1990). Furthermore, runs on individual banks should be distinguished from runs on the whole banking system. In the former case, deposits are redistributed from risky banks to sound banks (Kaufman 1988). Hence, bank runs are not necessarily bad and occasional runs can be optimal and better than without runs because they improve risk sharing (Alonso 1996).

Like bank runs, the bad impacts or third-party effects of bank failures are often exaggerated (Benston 1991). Some economists even go further to argue that bank failures do not differ from other firm failures since both generate negative externalities on their customers, although the mainstream view holds that banks are special because of information asymmetry (Fama 1985). While the macro-domino externalities of bank failures cannot be entirely denied, large-scale bank failures are

relatively rare events that do not warrant permanent costly government interventions such as deposit insurance, not to mention the empirical evidence indicating that bank runs and failures are an effect and not a cause of aggregate economic problems (Gorton 1986). Furthermore, the costs due to losses of information on borrowers can be avoided by absorbing failed banks with bridge banks or the purchase and assumption method. In brief, the purpose of banking regulation is not to prevent bank failures, let alone regulatory forbearance to allow insolvent banks to stay open simply because of regulators’ long-term career goals (Kane 1990).

While efficiency is undeniably an objective of regulation, competition and efficiency do not necessarily require a large number of firms in the industry, as in the case of contestability (Baumol et al 1982). Furthermore, regulators should not overlook the trade-off between dynamic efficiency and static efficiency. The former requires a small number of banks whereas the latter requires a large number (Allen & Gale 2000:ch. 8). By cross-subsidization, deposit insurance allows poor-quality banks to compete with prudent banks. But this does not imply higher efficiency, not to mention that it is unfair and increases systemic risk.

The positive impact of deposit insurance on financial development is also controversial. Besides the findings of Cull *et al.* mentioned earlier, countries with more extensive deposit insurance tend to have smaller capital and financial markets, reflecting its hindrance to non-bank financial development (Cecchetti & Krause 2000).

In sum, the effectiveness of deposit insurance in maintaining banking stability and its impacts on the safe and efficient development of the financial system depend

crucially on a country's institutional environment. Empirical evidence indicates that countries with weak infrastructure are exposing their banking systems to risk if they introduce deposit insurance (Demirgüç-Kunt & Kane 2002).

### **Further Debates**

Once introduced, deposit insurance is politically difficult to be repealed. Reform rather than abolition of existing systems is therefore a more promising solution to contain the perverse incentives from deposit insurance. As existing systems vary considerably in administration, funding, coverage, etc. (see Garcia 2000, Demirgüç-Kunt & Sobaci 2001, for details), we limit our discussion here to a few major directions of reform instead of the institutional details.

A solution to control bank risk is the risk-based capital adequacy requirements, originally introduced in 1987 and revised standards, commonly referred to as Basel II, have recently been adopted by many countries (see BIS 1999 for details). The essence is to limit banks' excessive risk-taking by requiring more risky banks to face higher costs of capital. Its effectiveness remains to be seen, as there are difficulties in applying market-value accounting and value-at-risk frameworks.

A similar measure is to impose risk-rated deposit insurance premiums. A higher premium is charged if a bank's portfolio is more risky. A traditional argument against risk-rated premiums is that they are either backward looking if they are not based on a thorough analysis of a bank's current asset portfolio or very costly if they are. More formally, Chan et al (1992) argue that fairly priced premium under asymmetric information is infeasible because of adverse selection and the time lag between audit and

premium adjustment, during which an undercapitalized bank may gamble for resurrection. However, Freixas and Rochet (1998) argue that fairly priced deposit insurance can be possible but undesirable under asymmetry information. Notwithstanding the theoretical debate, countries like the USA and Canada have recently introduced risk-rated premiums and initiated the global trend.

Other measures to control bank risk include lowering and limiting coverage and coinsurance requiring depositors of a failed bank to bear part of the accrued losses. These measures intend to minimize depositors' moral hazard and encourage them to impose market discipline on banks, as depositors tend to demand higher deposit interest rates or withdraw their deposits from risky banks. While coinsurance is still relatively rare, it has become more frequent among recently adopted deposit insurance schemes (Demirgüç-Kunt & Kane 2002).

Another measure to strengthen market discipline is privatization, for profit-maximizing private insurance companies have higher incentives than governments in monitoring banks. However, private insurers may lack the credibility to raise sufficient capital to support the huge volume of deposits nowadays and to cover losses from large-scale bank failures. In other words, they may be unable to contain contagious runs. By contrast, the government can maintain the solvency of the insurance fund and bolster by its power to tax (Mester 1992). Politically, closures of commercial banks may become public decisions that render difficulties to private schemes (Benston *et al.* 1986). These factors explain why deposit insurance schemes are predominantly administered by governments.

## **Empirical Evidence**

As deposit insurance have theoretically both good and bad outcomes, it is crucial to examine how it actually performs in practice. Our focus here is the empirical evidence of the effects of deposit insurance on banking stability and market discipline.

Due to data availability, early studies are mainly based on the US experience, some of which have already mentioned before (see also, e.g., Keeley 1990, Grossman 1992). Some other countries also share the perverse effect of deposit insurance. For example, Canada had enjoyed a long period of stability since 1923 without deposit insurance; and the introduction of deposit insurance in 1967 saw subsequent increases in failures and losses of insured institutions (Carr et al 1995).

A database furnished by the World Bank enables recent studies to address the empirical issues on a global scale. An econometric study based on 61 countries for 1980-97 indicates that deposit insurance on average tends to induce banking instability, particularly among countries with weak institutional environments (Demirgüç-Kunt & Detragiache 2002). Bank-level data for 43 countries over 1980-97 also show that deposit insurance weakens market discipline as it lowers banks' interest expense and makes it less sensitive to bank risk and liquidity. Similarly, market discipline is stronger in countries with stronger institutional environments and it can be enhanced by features such as coinsurance and private management of the insurance system (Demirgüç-Kunt & Huizinga 2000).

Like its theoretical counterpart, the empirical literature does not reach a consensus. Based on their findings, Eichengreen and Arteta (2002) conclude that

deposit insurance has favourable effects by providing protection from depositor panic but also unfavourable effects by weakening market discipline and destabilizing banking systems in emerging economies. Differences in the data and empirical techniques used are plausible sources of discrepancies in the findings. In addition, time can be a factor as deposit insurance is stabilizing in the short run but destabilizing in the long run (Chu 2003). Undoubtedly, more research is needed for a better understanding of the different impacts of deposit insurance.

## **Assessment**

The literature on deposit insurance is so voluminous that omissions of many important issues are inevitable (see FDIC 2000 for an overview of major studies). Overall, deposit insurance is a mixed blessing. On the one hand, it weakens market discipline and removes incentives for sound and prudent banking; on the other, it reduces banking instability -- but only when the regulatory infrastructure is strong. A key problem facing regulators is to identify the relevant factors affecting the optimal trade-off between stability and market discipline. The empirical and historical evidence indicates that there is no one-size-fits-all solution. Some countries are free from banking instability without deposit insurance, whereas some are repeatedly haunted by systemic crises even after the introduction of deposit insurance. As Kane (2000) correctly points out, country-specific factors should be taken into account when an explicit deposit insurance scheme is introduced. Simply put, do not just copy a system from another country.

## **Internet Sites**

Federal Deposit Insurance Corporation.

## Selected References

- Allen, F. and D. Gale. (2000) *Comparing Financial Systems*. Cambridge, M.A.: MIT Press.
- Alonso, I. (1996) "On Avoiding Bank Runs", *Journal of Monetary Economics* Volume 37, pp. 73-87.
- Basel Committee on Banking Supervision. (1999) *A New Adequacy Framework*, Basel: Bank for International Settlements.
- Baumol, W., J. Panzar, and R. Willig. (1982) *Contestable Markets and the Theory of Industrial Structure*. San Diego: Harcourt Brace Jovanovich.
- Benston, G. (1991) "Why Continue to Regulate Banks? An Historical Assessment of Federal Banking Regulations", pp. 434-49, in D. Chew (Editor), *New Developments in Commercial Banking*. Oxford: Basil Blackwell.
- Benston, G., R. Eisenbeis, P. Horvitz and E. Kane. (1986) *Perspectives on Safe and Sound Banking: Past, Present, and Future*, Cambridge, M.A.: MIT Press.
- Bernake, B. (1983) "Non-Monetary Effects of the Financial Crisis in the Propagation of the Great Depression", *American Economic Review*, Volume 73, pp. 257-76.
- Bordo, M. (1992) "The Lender of Last Resort: Alternative Views and Historical Experience", in A. Saunders, G. Udell, and L.J. White (Editors), *Bank Management and Regulation: A Book of Readings*. Mountain View, CA: Mayfield, pp. 43-52.
- Calomiris, C., and C. Kahn. (1991) "The Role of Demandable Debt in Structuring Optimal Banking Arrangements", *American Economic Review*, Volume 81, pp. 497-513.
- Carr, J., F. Mathewson, and N. Quigley. (1995) "Stability in the Absence of Deposit Insurance: the Canadian Banking System 1890-1966", *Journal of Money, Credit and Banking*, Volume 27, Number 113, pp. 7-58.
- Chan, Y.S., S.I. Greenbaum, and A.J. Thakor. (1992) "Is Fairly Priced Deposit Insurance Possible?" *Journal of Finance*, Volume 47, pp. 227-45.
- Cecchetti, S., and S. Krause. (2000) "Deposit Insurance and External Finance", mimeo, Ohio State University.
- Chu, K.H. (2000) "Saving Bonds as Deposit Insurance", *HKCER Letters* No. 62: pp. 6-8.
- Chu, K.H. (2003) "Deposit Insurance and Banking Crises in the Short and Long Run", in *Cato Journal*, Volume 23, Number 2, pp. 265-280.
- Cowen, T. and R. Kroszner. (1990) "Mutual Fund Banking: A Market Approach", *Cato Journal*, Volume 10, pp. 223-37.
- Cull, R., L. Senbet, and M. Sorge. (2002) "The Effect of Deposit Insurance on Financial Depth: A Cross-Country Analysis", *Quarterly Review of Economics and Finance*, Volume 42, pp. 673-94.
- Demirgüç-Kunt, A., and E. Detragiache. (2002) "Does Deposit Insurance Increase Banking System Stability?: An Empirical Investigation." *Journal of Monetary Economics*, Volume 49, pp. 1373-1406.
- Demirgüç-Kunt, A., and H. Huizinga (2000) "Market Discipline and Financial Safety Net Design", *Policy Research Working Paper 2430*. Washington, D.C.: World Bank.
- Demirgüç-Kunt, A., and E. Kane. (2002) "Deposit Insurance Around the Globe: Where Does it Work?", *Journal of*



- Economics Perspectives*, Volume 16, Number 2, pp. 175-95.
- Demirgüç-Kunt, A., and T. Sobaci. (2001) "Deposit Insurance Around the World: A Data Base", *World Bank Economic Review*, Volume 15, pp. 481-490.
- Diamond, D. (1984) "Financial Intermediation and Delegated Monitoring", *Review of Economic Studies*, Volume 51, pp. 393-414.
- Diamond, D., and P. Dybvig. (1983) "Bank Runs, Deposit Insurance, and Liquidity", *Journal of Political Economy*, Volume 91, pp. 401-19.
- Eichengreen, B., and C. Arteta (2002) "Banking Crises in Emerging Markets: Presumptions and Evidence", in M. Blejer and M. Skreb (Editors), *Financial Policies in Emerging Markets*, Cambridge, M.A.: MIT Press, pp. 47-94.
- Fama, E. (1985) "What's Different about Banks?" *Journal of Monetary Economics*, Volume 15, pp. 29-40.
- Federal Deposit Insurance Corporation. (2000) *Deposit Insurance: An Annotated Bibliography 1989-1999*. Washington D.C.: FDIC.
- Freixas, X., and J. Rochet. (1998) "Fairly Priced Deposit Insurance: Is it Possible? Yes. Is it Desirable? No", *Research in Economics*, Volume 52, pp. 217-32.
- Garcia, G. (2000) *Deposit Insurance: Actual and Good Practice*. IMF Occasional Paper No. 197, Washington DC: International Monetary Fund.
- Golembe, C. (1960) "The Deposit Insurance Legislation of 1933: An Examination of Its Antecedents and Its Purposes", *Political Science Quarterly*, Volume 75, pp. 181-200.
- Gorton, G. (1988) "Banking Panics and Business Cycles", *Oxford Economic Papers*, Volume 40, pp. 751-81.
- Grossman, R. (1992) "Deposit Insurance, Regulation, and Moral Hazard in the Thrift Industry: Evidence from the 1930s", *American Economic Review*, Volume 82, pp. 800-21.
- Kane, E. (1985) *The Gathering Crisis in Federal Deposit Insurance*. Cambridge, M.A.: MIT Press.
- Kane, E. (1989) *The S&L Insurance Mess: How Did it Happen?* Washington, D.C.: Urban Institute.
- Kane, E. (1990) "Principal Agent Problems in S&L Salvage", *Journal of Finance*, Volume 45, pp. 755-64.
- Kane, E. (2000) "Designing Financial Safety Nets to Fit Country Circumstances", *Policy Research Working Paper 2453*, Washington, D.C.: World Bank.
- Kaufman, G.G. (1996) "Bank Failures, Systemic Risk, and Bank Regulation", *Cato Journal*, Volume 16, pp. 17- 45.
- Kaufman, G.G. (2000) "How Real is the Risk of Massive Banking Collapse? Evaluating the Latest Methods Proposed to Maintain Stability in International Financial Markets and to Minimize Systemic Risk", *Regulation*, Volume 23, pp. 25-30.
- Keeley, M. (1990) "Deposit Insurance, Risk, and Market Power in Banking", *American Economic Review*, Volume 80, pp. 1183-1200.
- Levine, R., N. Loayza and T. Beck. (2000) "Financial Intermediation and Growth: Causality and Causes", *Journal of Monetary Economics*, Volume 46, pp. 31-77.
- Litan, R. (1987) *What Should Banks Do?* Washington, DC: Brookings Institution.
- Merton, R. (1977) "An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees: An Application of Modern



- Option Pricing Theory”, *Journal of Banking and Finance*, Volume 1, pp. 3-11.
- Mester, L. (1992) “Curing Our Ailing Deposit-Insurance Scheme”, in A. Saunders, G. Udell, and L.J. White (Editors *Management and Regulation: A Book of Readings*, Mountain View, CA: Mayfield, pp. 383-94.
- Sprague, I. (1986) *Bailout: An Insider Account of Bank Failures and Rescues*. New York: Basic Books.
- World Bank (2001) *Finance for Growth: Policy Choices in a Volatile World*. Washington, DC: World Bank.

*Kam Hon Chu*  
*Memorial University of Newfoundland*  
*St. John's, Newfoundland & Labrador*  
*Canada*  
*kchu@mun.ca*

## Development Governance

*Richard Peet*

### **Introduction**

Development governance refers to the institutions that regulate development and the policies employed by them. In the post-World War II period, under the Bretton Woods system, global governance organizations took increasing control over the development policies employed by national governments. While “global” in terms of sphere of operations, development governance institutions cluster their headquarters in a few places. The resulting centers of global governance power are exclusively cities in Western capitalist countries: either the capitals of leading national powers, overwhelmingly New York (for example, the United Nations) and Washington DC (the International Monetary Fund and World Bank), or definitely Western, but officially neutral, political spaces, as with Geneva, Switzerland (the World Trade Organization and many U.N. aid and development agencies).

“Governance” means regulation, direction, management and continuing control over development policy by institutions that are, at best, only indirectly elected, as compared with regulation of national policy by democratically elected governments. In governance, power passes from electorates *through* governments to their representatives in governance institutions. As a result development governance suffers from problems of legitimacy, especially in that the people being developed have little recourse to the institutions performing the governance. Under governance, policy is legitimated as true, correct and socially just by science and

expert opinion, rather than by elections and democratic consent. Based on this, critics have argued that development governance may be read as a new, kinder-faced version of Western imperialistic control, this time using the optimism embedded in development as a more effective, less obviously violent means (eg. Escobar 1995).

On the surface, the position increasingly taken by development governance institutions is that the policies they advocate—debt relief, for instance—express concern for the world’s poorest people. Yet development governance institutions condition aid to poor countries on their adoption of a set of policies that insist on an exaggeratedly market-organized and export-oriented economy—policies referred to, particularly by critics, as “neoliberal” (Peet et al 2003; Harvey 2005). This set of neoliberal policies is said to produce the economic growth that creates the jobs that reduce poverty. Yet these are also policies that favor free enterprise, free trade, unrestricted private capital investment, and similar measures. And the only reason for private investment in developing countries is to make profit. Development governance institutions therefore act, at least in part, in the interest of global capital. Looking from a critical viewpoint, therefore, development governance serves Western, capitalist interests while pretending concern for the Third World poor. From that realization springs such contention that massive protest greets just about every meeting of the International Monetary Fund, the World Trade Organisation, and other development governance institutions.

### **Policy Regimes**

The global capitalist economy is made up from millions of specialized production

units, each motivated by the competitive drive for profit, tied together organizationally by webs of markets. Theoretically, in neoclassical economics, governmental intervention into market mechanisms is unnecessary and, in extreme neo-liberal versions of economic thought, state intervention is harmful to economic efficiency and dangerous to political liberty (Hayek 1956). But practically, market systems have always been embedded within institutional contexts that include state legislation, economic policy, governmental intervention and, increasingly, governance regulation. Since the late nineteenth century, the emerging global capitalist system has been governed under three political-economic policy regimes: Liberal Democracy, predominating between 1870 and 1930; Keynesian Democracy, predominating between 1945 and 1973; and Neoliberal Democracy predominating between 1980 and the present; the years 1930-45 and 1973-80 represent transitional periods when regimes contended for dominance.

The original, liberal democratic regime entailed minimal state intervention into the domestic economies of the industrial countries, yet strong, political intervention overseas, especially in the Third World, under what is termed “imperialism”. The Keynesian policy regime was characterized by counter-cyclical macro-economic management by an interventionist state committed to achieving full employment and higher incomes for everyone. This regime responded to the Depression of the 1930s by employing an interventionist state to stabilize economic development, manage demand, and democratize social benefits. Regional differences in political-economic traditions informed three main variants:

Social Democratic Keynesianism in Western European countries and their former settler colonies; Liberal Democratic Keynesianism in the US; and Developmental State Keynesianism in Japan and many of the newly industrializing countries of Latin America and East Asia.

The convention is that Keynesianism entered into crisis in the 1970s characterized by stagflation—that is, high rates of inflation coinciding with high rates of unemployment. Its successor, the Neoliberal Democratic policy regime, revives late nineteenth century Liberalism by partially withdrawing the nation state from macroeconomic management (within an upward displacement of macro-power to global governance institutions) and with an increased reliance on market mechanisms, with free trade and freedom of capital to invest where it will. Neoliberalism employs monetarist economics under the conception that macroeconomic problems like inflation and debt derive from excessive government spending. While regional variations in speed of adoption, and level of commitment, persist, this regime responded positively to the globalization of economy, society and culture of the late twentieth and early twenty-first centuries and indeed helped to organize the new world capitalist system.

When market systems were primarily national in geographic scope, the necessary governance framework could similarly be limited to the nation state. But the globalization of capitalism has long required and, indeed has been led by, the parallel globalization of institutional regulation. In the first globalization of the late nineteenth century, British national institutions precariously managed inter-national economic relations, with sterling as agreed-upon currency of account, and the City of

London as the global, hegemonic center of political-economic power. Pax Britannica was challenged during two World Wars, with Britain demoted in their aftermath to the status of sub-hegemony—although remnants of the previous hegemony remain, for example in London's strong position in global currency exchange and the international bond market. In the second globalization of the late twentieth century, American institutions precariously manage inter-national economic relations, with the US dollar as currency of account, Wall Street, New York City as the main center of political-economic power, and Washington DC as corresponding center of political power. However, the second globalization learns from the collapse of the first, in the Depression of the 1930s, that nationally-based institutions cannot suffice, no matter how powerful Pax Americana may seem. Hence the IMF, World Bank and WTO, the European Union, the G7 and G8 using, as well as US dollars, the Yen, Deutschmark, Pound and Euro in a spectacular array of regional, supra-national, inter-national and truly global institutions, greatly influenced, but only partly disciplined, by American political-economic power.

### **Origins of Development Governance**

While the world was still engaged in World War II, forty-four nations, led by the US and the UK, met at Bretton Woods, New Hampshire, in the United States, between July 1 and 22, 1944 to discuss economic plans for the post-war peace. The very idea of governing the international economy through global institutions was made possible by the competitive anarchy of the inter-war period. In reaction, governments sought to secure world peace and prosperity through inter-national economic

cooperation. Such cooperation would be based on a world market in which capital and goods might move freely, while being regulated by global institutions operating in the general interests of greater stability and predictability. Three governance institutions were envisaged: the International Monetary Fund (IMF), the International Bank for Reconstruction and Development (IBRD, later known as the World Bank), and an International Trade Organization (ITO) that came into being only as the General Agreement on Trade and Tariffs (GATT), but much later became the WTO. The IMF and IBRD were formalized as organizations during the Bretton Woods conference, while the proposal for an ITO was part of a separate Havana Charter of 1947.

At Bretton Woods discussion about the formation of international organizations dwelt almost exclusively on the IMF. The IBRD was a mere afterthought, and what little discussion occurred centered exclusively on its possible role in the post-war reconstruction of Western Europe. On the few occasions that the development of poor countries was, briefly, mentioned, issues like poverty never came up. Indeed such were the pre-occupations of the Europeans and Americans at the time, that labels like “poor countries,” or the more critical term “underdeveloped counties,” did not exist as functional geographical categories – rather countries outside Europe and North America were referred to as “the colonies.”

The idea of using the World Bank, especially, in the development of Third World countries emerged slowly during the 1950s and 1960s. Now, however, the World Bank operates as a development governance agency with a mission statement that proclaims “our dream is a world without

poverty.” For this the World Bank Group lends some US\$17 billion a year to client countries. What is more, it sets the conditions under which further billions, in loans and grants, flow to Third World and post-Communist countries. Taken together, these two interventions, direct loans and the setting of policy conditions, make the World Bank the most important development institution in the world.

The main governance institutions, the IMF and the World Bank, have become powerful institutions. At any time the economies of 120 nation states, and the livelihoods of 2.5 billion people, might be under their direct supervision, within what used to be called, and is still best termed, “structural adjustment”, but now is more charitably termed “debt relief”, “poverty alleviation” and meeting “Millennium Goals”.

### **Evolution of Development Governance Policy**

In their early days of dealing with Third World countries, in the 1950s and early 1960s, development governance institutions limited themselves to “project lending” – that is, long-term loans for infrastructure like dams, electricity, communications and transportation systems in line with what little economic theory at the time dealt with developmental issues. The idea of early development economics was to remove blockages to economic growth, or set the preconditions for growth, by making capital investments (project lending) that would raise the general level of national productivity.

However, as the 1950s progressed, the World Bank, along with other development institutions, evidenced a rising interest in broader issues of income distribution and

poverty. This sudden expression of concern for the masses of poor people in the world stemmed from Cold War engagement between the US and the USSR, with the Third World as ideological battle ground. The Eisenhower administration moved in the direction of increasing aid to Third World countries in the late 1950s by arguing that aid was “an investment for peace.” But the Robert McNamara period as president of the World Bank (1968-81) saw major changes. The period 1968-73 saw a number of temporary enthusiasms in development governance—including population control, employment policy, nutrition, health, water supply—all culminating in the later 1970s with a “basic needs approach” to lending. That is, giving greater attention to need rather than output in the allocation of resources.

After several miss-starts, the final form of development was called “integrated rural development projects.” Bank loans and technical assistance for key agricultural inputs (fertilizer, seeds) and infrastructure (water, electricity, roads) aimed at small farmers in a specific geographical area (hence a “project”), that would increase productivity, providing surpluses that could be sold in markets, which in turn would raise (money) incomes. However, such projects proved to be easier to outline on paper in Washington than carry out in practice in the field—even the World Bank’s own Operations Evaluation Department called most such projects “failures.” Well before the end of McNamara’s presidency, the ardor had gone from the poverty initiative which, at its height, had amounted to no more than a third of Bank lending, with the rest going to the traditional areas of infrastructure investment. Debt and balance of payments in Third World countries

became leading issues, with “structural adjustment” as solution (Kapur 1997).

### **Structural Adjustment**

Dissatisfied with the results of the basic needs approach, McNamara said in 1979 that the World Bank should use program loans to induce “reforms” in recipient, mainly middle-income, countries—“reform” meaning structural adjustment lending to promote export-orientation and trade liberalization. The World Bank followed the lead of its more senior partner, the IMF, under a division of labor that allocated “stabilization programs” (short-term adjustment lending) to the IMF and longer term “structural adjustment lending”, aimed at correcting deeper “structural” problems, to the World Bank. In 1980 the World Bank laid out the general conditions under which structural adjustment loans would be made available. The argument was that the new conditions faced by Third World countries—deteriorating terms of trade and growing current account deficits—forced them to reconsider how they might “adjust” their development patterns and economic structures. The World Bank’s new lending program would provide loans that were policy-based, rather than project-based, extended over several years, and would provide direct support for specific policy reforms decided upon during “dialogue” with the borrowing country.

This turn to a more policy-interventionist stance, evident in the late McNamara years, was re-enforced in the early 1980s after Margaret Thatcher became Prime Minister of Great Britain in 1979, Ronald Reagan was elected President of the United States in 1980, and Helmut Kohl was elected Chancellor of West Germany in 1982, and with a change in the World Bank’s

presidency in 1981 to A.W. Clauson, formerly head of the Bank of America. Under the Clauson presidency poverty took a back seat to new driving forces of macroeconomic policy, stabilization and balance of payments adjustments, all understood within a more right-wing doctrine of the strict limits of governmental intervention and the virtues of flexible, self-adjusting, free markets. Structural adjustment became the main means of carrying these political beliefs into economic practice—structural adjustment loans focused on macroeconomic policies and institutional change at the country level and for the medium term.

By the middle 1980s, a new set of development governance policies was generally agreed on under what Williamson (1990, 1997) termed a “Washington Consensus. By “Washington” Williamson meant the political Washington of the US Congress and senior members of the administration, and the technocratic Washington of the international financial institutions, the economic agencies of the US government, the Federal Reserve Board, and the think tanks, as the numerous policy research and advocacy institutions congregating in capital cities are termed. In brief, said Williamson (1990: 18), the economic position Washington concurred on in setting policy for the rest of the world, but did not necessarily follow itself, could be summarized as “prudent macroeconomic policies, outward orientation, and free market capitalism.”

### **Debt Relief**

Beginning with Mexico’s default, in 1982-4, on payments on its foreign obligations, a Third World debt crisis soon became the issue of paramount importance in

development governance. At first, an assumed short-term liquidity crisis in middle income Third World countries was addressed by IMF stabilization programs. Debt relief took the form of payment rescheduling, sometimes on concessionary (low interest) terms, sometimes coupled with new loans. Creditor governments formed a committee to deal with debt relief, in consultation with the IMF, was hosted by the French Treasury Department, and known as the “Paris Club”. Repeated Paris Club re-scheduling of debts led official lenders eventually to recognize that a new approach was needed for these countries.

Continuing efforts by the US to find a position adequate for responding to a deteriorating debt situation culminated in a proposal made in October, 1985 by James A. Baker III, Secretary of the Treasury between 1985 and 1988 in the Reagan administration, in what came to be known as the “Baker Plan.” The idea was that the IMF and the World Bank should join forces to increase the amount of loans available from both institutions and the commercial banks. But loans would be made conditional on “policy improvements in the macroeconomic framework” under structural adjustment programs—the “policy improvements” being in line with neoliberal ideas about the causes of economic growth (markets, privatization, de-regulation of private enterprise, reducing state deficits and so on). However, over the next three years the official view became that the Baker plan had failed, and that more extreme debt-reduction measures were necessary. When the (Republican) Bush administration assumed office in 1989, the new Secretary of the Treasury, Nicholas Brady, announced that the only way to address the debt crisis was to “encourage” the banks to engage in

“voluntary” debt-reduction. Under what became known as the Brady Plan, countries were to implement market oriented structural adjustment, as with the Baker Plan, but this time in exchange for a reduction of commercial bank debt and, often, new loans from commercial banks and multilateral lending agencies. By 1994, 18 countries had agreed to “Brady deals” forgiving \$60 billion of debt and covering about \$190 billion in bank claims. The typical Brady deal leads to 30 to 35 percent forgiveness of a country’s debt. Behind this flurry of debt rescheduling and relief activity, strong political pressure was exerted by national governments, especially the US, usually operating in concert with the Paris Club and the IMF. The main concern of this group of actors was preserving the banking system in the face of the possibility of repudiations of hundreds of billions of dollars in un-payable debts.

From 1986 onwards, the World Bank and other development governance institutions came under increasing criticism from Non Governmental Organizations (NGOs) representing social and environmental movements, and from many other concerns that mobilized people and ideas and motivated liberal political constituencies in donor countries. Widespread and increasingly well-informed criticism of the Bretton Woods institutions was led by Jubilee 2000, a faith based coalition that believed that the 2000<sup>th</sup> anniversary of the birth of Christ was an opportune time for a “jubilee” in the form of a forgiving of debts. Conventional, and even conservative, opinion had also, by the late 1980s, reached the conclusion that debt levels were unsustainable, and that some kind of organized relief was necessary. While the debt crises of the 1980s mainly concerned

middle-income Third World countries, like Mexico, Brazil and Argentina, in the 1990s the main emphasis of IMF/World Bank debt management shifted towards the lowest income, Third World countries. Loans to these very poor countries had been made mainly from official sources, as with government to government loans, export credits, official development assistance and loans from the IMF, World Bank and regional development banks (Birdsall and Williamson 2002). Responding to concern expressed by developed countries, the IMF and World Bank began their Heavily Indebted Poor Countries (HIPC) initiative in 1996.

The HIPC Initiative was intended to manage, and even “resolve” in the IMF’s optimistic language, the debt problems of the most heavily-indebted poor countries (originally 41 countries, mostly in Africa) with total debt of about \$200 billion. In these countries debt service obligations consumed large parts of countries’ export earnings. Half of the 615 million people in the current HIPC countries live on less than \$1 a day. The HIPC initiative, in the IMF’s (2000) words, “seeks a permanent solution to these countries’ debt problems by combining substantial debt reduction with policy reforms to raise long-term growth and reduce poverty.” By adopting policies judged “sound by the international community,” debt relief to the eventual extent of \$60 billion would be granted. The HIPC initiative was “enhanced” in 1999 to provide interim debt relief that immediately reduced debt service costs. The enhanced facility joined debt relief more obviously with poverty reduction.

To qualify for assistance under the HIPC initiative, or to get concessionary loans from the IMF or World Bank, countries have to

prepare comprehensive Poverty Reduction Strategy Papers (PRSPs) with the participation of members of civil society. However, a number of studies have critically examined the new initiatives. An extensive investigation by the (London-based) World Development Movement, found civil society groups unsatisfied with the extent of public involvement in drawing up the strategy papers. The report said that the policy content of the new strategies does not constitute a major change from the past. Although the rhetoric may be poverty-focused, the actual policies do not have clear poverty reducing consequences. The strategies still focus on economic growth without, on the most part, addressing how this growth is redistributed to the poor. Indeed, the core macroeconomic elements have changed little from the old structural adjustment programs, with continued adherence to privatization, liberalization and a reduced role for the state (Marshall & Woodroffe 2001).

### **Poverty Reduction**

Some commentators find the World Bank shifting at the end of the 1980s and in the early 1990s to a revised neoliberal model stressing market-friendly state intervention and good governance (political pluralism, accountability and the rule of law) with a renewed emphasis on social issues like poverty and education and a dedication to debt reduction (Kiely 1998). Thus, in the 1990s, various *World Bank Development Reports* outlined a new “holistic approach” to development involving social safety nets, poverty, health, education, environment, rural areas and gender considerations, in concert with conventionally neoliberal areas, like increased property rights, trade liberalization and privatization. In 1997



James D. Wolfensohn, who became president of the World Bank in 1995, called for closer relations with other institutions and civil society actors to increase the effectiveness of development governance. In 1998, the World Bank carried out a series of consultations with government representatives, bilateral donor agencies, multilateral financial institutions, academics, NGOs and other civil society organizations, together with the private sector. Also in 1998 Wolfensohn, suggested an integrated approach to development based on a framework articulated and “owned” by the country itself, aimed at poverty reduction and sustainable development, known as the Comprehensive Development Framework (CDF). In many respects, this is the World Bank’s policy answer to criticisms that structural adjustment does little to alleviate poverty, and indeed causes it.

As outlined in the 1999/2000 *World Development Report*, the CDF has two complementary parts: a stable macroeconomy shaped by “prudent fiscal and monetary policies;” and the CDF itself, stressing honest governments, strong property and personal rights, supported by an efficient legal and judicial system, human development, as with education and health, physical infrastructure, and sectoral elements like integrated rural development strategies and urban management (World Bank 1999). The CDF is based on four principles designed to guide implementation of future development strategies in specific countries. First, the framework emphasizes the need for a long-term, holistic vision of development that considers structural and social issues simultaneously, as well as acknowledging the inter-linkages between all economic sectors. Second, the country receiving assistance or loans needs to be in

control, taking ownership of the process. This entails the respective governments building consensus domestically and consulting with as many different civil society and private sector actors as possible. Third, the necessary prerequisite to country ownership is an emphasis on partnership between the “stakeholders” in the development process, again incorporating government, civil society and the private sector of the country concerned, and the main external actors, such as the World Bank, donor countries and bilateral aid organizations. Finally, the CDF calls for regular assessments of actual development outcomes to ensure accountability towards meeting the goals set out in each country’s long-term vision (World Bank 1999; 2000a, 2000b).

The CDF was part of a larger effort by the World Bank to produce what it terms a new consensus in the international development community, comprised of donor organizations and bilateral and multilateral agencies, such as the UN and the OECD, on the ingredients for successful development policies. The World Bank was trying to reposition itself within a much larger group of international organizations that emphasize a dedication to alleviating poverty and providing development assistance. Emphasizing agreement on the fundamentals of development policies with other institutions of the UN, which historically have been much better received by the Third World and NGOs, the World Bank signaled a shift in approach to its many critics. The World Bank now says that participation, including poor or marginalized groups and the private sector is crucial to building country ownership of national development strategies financed by the Bank.

However the turn of the century saw a broadening and deepening of the discussion within and around development governance institutions on issues of global poverty. A resolution adopted by the General Assembly of the United Nations in September 2000, titled the UN Millennium Declaration said that the central challenge its members faced today was to ensure that globalization becomes a positive force for the world's people in a context in which benefits were unevenly shared and costs unevenly distributed. Only through broad and sustained efforts to create a shared future, based upon a common humanity, in all its diversity, could globalization be made inclusive and equitable. These efforts included policies and measures corresponding to the needs of developing countries and economies in transition that were formulated and implemented with their effective participation. The industrialized countries were called on: to adopt policies of free access for exports from the least developed countries; to implement an enhanced program of debt relief for the heavily indebted poor countries and to cancel all official bilateral debts of those countries in return for their making commitments to poverty reduction; and to grant more generous development assistance, especially to countries that were genuinely attempting poverty reduction.

The Millennium Development Goals included: halving, by the year 2015, the proportion of the world's people whose income was less than one dollar a day and the proportion of people suffering from hunger and unsafe drinking water; ensuring, by the same date, that children would be able to complete a full course of primary schooling and that girls and boys would have equal access to all levels of education;

reducing maternal mortality by three quarters, and under-five child mortality by two thirds, of their current rates; and reversing the spread of HIV/AIDS, malaria and other major diseases afflicting humanity (UN 2000). Several rounds of negotiations and conferences followed, essentially re-affirming country commitments to reaching these goals. At their meeting in early June 2005, the finance ministers of the G7 countries agreed to provide additional financial resources, ensuring that the financing capacity of the World Bank, the IMF and the African Development Bank was not reduced by the HIPC initiative. This would eventually lead to 100 per cent debt cancellation of outstanding obligations of 18 of the poorest countries in the world. The agreement was then formalized later that year at a G8 summit meeting in Gleneagles, Scotland, after widespread publicity including "Live 8" rock concerts headed by singers Bob Geldoff and Bono. The idea was to contribute to achieving the UN Millennium Development Goal of halving world poverty by 2015. This "writing off \$40 billion owed to international agencies" mainly amounted to a re-financing of the IMF and World Bank's HIPC initiative. Point 2 of the G7 declaration reads as follows:

"We reaffirm our view that in order to make progress on social and economic development, it is essential that developing countries put in place the policies for economic growth, sustainable development and poverty reduction: sound, accountable and transparent institutions and policies; macroeconomic stability; the increased fiscal transparency essential to tackle corruption, boost private sector development, and attract investment; a

credible legal framework; and the elimination of impediments to private investment, both domestic and foreign.”

The aspect of point 2 seized on by the media was “good government practices” as with transparency, anti-corruption and credible legal frameworks. The other aspect of point 2, macroeconomic stability, private sector development, and removing impediments to private investment, domestic and foreign, together with sections on free trade and open markets in later points, went unmentioned. But it was exactly the power relations involved with development governance institutions telling poor countries how they must run their economies, if they wanted to receive debt relief, that had brought hundreds of thousands of demonstrators onto the streets in protest whenever the World Bank and the IMF met over the preceding decade. The whole issue of development governance and policy continued to be awash with controversy.

### **End of Poverty?**

Perhaps the leading book appearing at the time, when great issues of global poverty, debt and development re-appeared as specters troubling global consciousness, was written by Jeffrey D. Sachs (2005), the world’s most famous development economist. The work was intended as a handbook on “how we could be the first generation to outlaw the kind of extreme, stupid poverty that sees a child die of hunger in a world of plenty”. The main argument of the book was that “the key to ending extreme poverty is to enable the poorest of the poor to get their foot on the ladder of development” (Sachs 2005:244). The extreme poor, Sachs argued, lack six major kinds of capital: enough human capital, as

health, nutrition and skills, to be economically productive; business capital, as machinery and transport, to increase productivity; infrastructure that forms critical inputs into business productivity; natural capital that provides the environmental services needed by human society; public institutional capital that underpins peaceful and prosperous division of labor; and knowledge capital that raises productivity and promotes physical and natural capital.

Breaking the poverty trap involved donor-based investments that raise the level of capital per person, producing a capital stock high enough that the economy is sufficiently productive to meet basic needs. Without outside donor funds the necessary investments simply cannot be financed. Thus, ending global poverty by 2025 required a global compact between rich and poor countries, as with the UN Millennium Project, whereby the rich countries follow through on previous pledges to provide 0.7% of their GNP as development aid. Indeed the bottom line was about \$135-\$195 billion a year in assistance, significantly less than the 0.7% figure. As Sachs said: “The point is that the Millennium Development Goals can be financed within the bounds of the official development assistance that the donor countries have already promised” (Sachs 2005:299). This essentially means that the US, presently contributing 0.15% of GNP as development assistance, would have to contribute half the necessary increase, followed by Japan, Germany, France, Italy and Britain. Hence, our generation, heir to two and a half centuries of economic progress, could realistically envision a world without extreme poverty.

But why should the people of the rich countries of the world give this kind of

money? The answer, for Sachs (2005:331), was that “hard evidence has established strong linkages between extreme poverty abroad and the threats to national security”. An economy stuck in the poverty trap often leads to state failure and failed states are seedbeds for violence, terrorism, drug trafficking and disease. If the US and Western Europe wanted to spend less time responding to failed states in the post 9/11 era, they would have to reduce the number of failed economies. It had been done before, with the Marshall Plan, meant to ensure Europe’s economic stability and strategic security in the post-war era. At the Rio Summit on Sustainable Development and with the Monterrey Consensus, the developed countries committed to doing so again. The richest of the rich should therefore come through with their contributions as a “profound and meaningful demonstration of our generation’s unique moment to secure global well-being” (Sachs 2005:346).

Assuming that declarations by the UN and rich-country commitments make a difference, as with debt relief, this argument seemed persuasive. But there are problems even with this generous statement. One is the ability of charity-based theories to understand the causes of world poverty and thus suggest policies that might indeed enable the end of poverty. The argument is based on a largely conventional economic-historical geography, essentially following Rostow’s (1960) notions of take-offs, ladders of development and successful transformations to affluence. Yet under all existing aid and debt relief schemes, to get their money, poor countries have to agree to open their markets to foreign competition, privatize public enterprises, withdraw the state from service provision, reduce state

budget deficits, re-orient their economies to export-orientation, flexibilize their labor markets, and so on down a list written under the belief that markets and free competition can guide any economy into the magic realm of growth, up the ladder of development in Sachs’ terms, if only workers are made more employable. But opening markets, as Sachs suggested, means losing protected jobs—that is, creating unemployment in the name of “efficiency” in countries where skilled and educated labor is already under-used.

Privatization meant introducing the profit motive into, for example, water or electricity supply, and cutting off anyone who cannot or will not pay the higher rates. Reducing swollen state deficits in the name of fiscal responsibility may sound fine until it is remembered that desperate people often rely for their lives on state-supported food subsidies and free health care. As for exporting more, the problem is “export what?” China now monopolizes cheap labor manufacturing and the prices of coffee, cocoa and cotton are volatile and declining in the long run, so that already-productive small farmers break their backs working for less than nothing. At the same time local food production is undercut by focusing on export crops and removing tariff protections, producing a dangerous mass vulnerability to episodes of starvation. Labor market flexibility meant attacking the unions, paying lower wages and eliminating what few laws might exist to protect workers. And finally, even the desired outcome, economic growth, does not necessarily reduce poverty, especially when growth follows the neoliberal policy design. In other words, to earn “aid”, supplicant countries have to restructure their economies neoliberally, so that they reward foreign investment.

The ethical argument presented by Sachs to encourage rich countries to increase development aid poses additional problems. After arguing that the aim of assistance was to make people healthier for their own sakes, the book concludes that the rich countries should invest in poor countries for the sake of their own security, to prevent failed states, and to prevent the poor from becoming terrorists. Rich countries should invest in poor countries because they fear them. This is part of a tendency evident after the attack on the World Trade center in 2001 to link issues of development with issues of national security—for example, the President of the World Bank appointed in 2005, was Paul Wolfowitz, a prominent neoconservative who had played a major role in planning the US-led invasion of Iraq in 2003. By comparison, critics argue, the only viable reason for aid might be re-stated as reparation for damages done by the First World countries in the past, and continuing into the present. In other words, the will to assist the Third World has to come from a sense of global justice, from the critical understanding that the wealth of a few caused the poverty of the many, from anti-imperialist sentiment rather than benevolence. These are some of the ongoing debates surrounding one of the most contentious issues of the early twenty-first century: development governance.

### Internet Sites

Millennium Development Goals.  
[www.un.org/millenniumgoals/](http://www.un.org/millenniumgoals/)  
 United Nations development Program.  
[www.undp.org](http://www.undp.org)  
 World Bank. [www.worldbank.org/](http://www.worldbank.org/)

### Selected References

Birdsall, N. and J. Williamson (2002)  
*Delivering on Debt Relief: From IMF*

*Gold to a New Aid Architecture.*  
 Washington: Center for Global Development.

Escobar, A. (1995) *Encountering Development: The Making and Unmaking of the Third World.* Princeton: Princeton University Press.

Harvey, D. (2005) *A Short History of Neoliberalism.* New York: Oxford University Press.

Hayek, F. (1956) *The Road to Serfdom.* Chicago: University of Chicago Press.

IMF. (2000) *The Logic of Debt Relief for the Poorest Countries.* Washington: IMF.

Kapur, D., J.P. Lewis and R. Webb. (1997) *The World Bank: Its First Half Century.* Washington: Brookings Institution Press. 2 vols.

Kiely, R. (1998) "Neo-Liberalism Revised? A Critical Account of World Bank Concepts of Good Governance and Market-Friendly Intervention" *Capital and Class*, Number 64: 63-88.

Marshall, A. and J. Woodroffe. (2001) *Policies to Roll-back the State and Privatised? Poverty Reduction Strategy Papers Investigated.* London: World Development Movement.

Peet, R. et al. (2003) *Unholy Trinity: The IMF, World Bank and WTO.* London: Zed Press.

Rostow, W.W. (1960) *The Stages of Economic Growth: A Non-Communist Manifesto.* Cambridge: Cambridge University Press.

Sachs, J.D. (2005) *The End of Poverty: Economic Possibilities for Our Time.* Foreword by Bono. New York: Penguin Press.

United Nations (2000) *Resolution Adopted by the General Assembly: 55/2 United Nations Millennium Declaration.* UN

- Williamson, J. ed. (1990) *Latin American Adjustment: How Much has Happened?* Washington: Institute for International Economics.
- Williamson, J. (1997) "The Washington Consensus Revisited" in L. Emmerij ed *Economic and Social Development into the XXI Century* Washington: Inter-American Development Bank, pp. 48-61.
- World Bank (1999) *World Development Report*. New York: Oxford University Press.
- World Bank (2000a) Background and Overview of the Comprehensive Development Framework. Washington DC: World Bank.
- World Bank (2000b) *Comprehensive Development Framework. Country Experience*. Washington DC: World Bank.

*Richard Peet*  
*Graduate School of Geography*  
*Clark University*  
*Worcester, Massachusetts, USA*  
[rpeet@clarku.edu](mailto:rpeet@clarku.edu)

## Economic and Financial Crises

*Christian E. Weller*

### **Introduction**

The ‘global subprime crisis’ of 2007-2009 was seen as the worst global financial-economic crisis since the Great Depression of the 1930s. It occurred as the world has become more integrated through barriers to capital movements being eliminated. Liberalization of financial markets, which comprises both a deregulation of local financial markets and the elimination of controls on external capital flows, is based on a theoretical framework, known as financial liberalization. This theory holds that fewer regulations on financial capital are an integral part of a successful development strategy. Allowing capital to move freely would supposedly allow investors to seek out the most productive investments, which in turn should result in faster growth for the recipient economies than otherwise would be the case.

The benign view of financial liberalization, though, has been called into question as evidence has mounted that greater deregulation has directly contributed to the increased frequency of crises. In a comprehensive study of banking crises, the IMF found that during the period from 1980 to 1996, when liberalization became increasingly popular, two thirds of the IMF’s member countries had undergone significant banking sector problems. Increasingly, researchers have linked the two phenomena, liberalization and crises as liberalization tends to give rise to opportunities for speculative financing.

Financial crises have repercussions for crisis countries and their trading partners. Financial sector problems, be they troubled

banks or falling exchange rate values, can quickly spell trouble for the afflicted economies, lowering output, increasing unemployment and raising poverty. Moreover, the impact of a crisis is often not restricted to one economy, but it can spread to other economies through international trade and capital linkages. Hence, the challenge for policymakers is to ensure financial stability, while making sufficient funds available for productive investments. Thus, much of the debate has focused on potentially stabilizing institutions.

### **Financial and Economic Crises**

Before considering the causes of and possible remedies for increasing financial and economic instabilities, it is useful to delineate some definitions. Financial crises can generally be understood as a sharp deterioration of a group of financial indicators, such as interest rates or asset prices—including foreign exchange rates. Hence, financial sector crises typically come in two forms, banking crises or currency crises. Although there is no fast and hard rule what constitutes a banking crisis, most researchers consider “bank runs or other substantial portfolio shifts, collapses of financial firms, or massive government intervention” as crisis (Lindgren et al. 1996:20). Noticeable problems short of a crisis are considered significant banking problems, but this often proves too broad a measure to be useful for economic analysis.

In comparison, currency crises, or balance-of-payment crises reflecting the main underlying cause of the crisis, are reflected by a rapid devaluation of the currency. However, a country may be able to fend off a speculative attack on its currency by selling its reserves or by raising interest rates (Eichengreen et al. 1995). Typically

researchers have used an index created by calculating a weighted average of exchange rate changes, changes in official reserves and interest rate movements to measure speculative attacks or currency crisis. If the index increases more than a pre-set amount, generally two standard deviations, it is considered a crisis or a speculative attack (Eichengreen et al. 1995). Due to measurement problems, researchers have moved towards creating this index, though, by only including changes of exchange rates and of official reserves (Kaminsky & Reinhart 1999).

Often banking and currency crises are highly correlated, with one leading the way for the other. Moreover, financial sector problems tend to spill-over into the real sector, most directly through credit crunches or imported inflation, thereby depressing domestic economic activity. But there is no consensus on what constitutes a macro economic crisis in the empirical literature. Some have used rescheduling of external debt, arrears on external payments or inflation rates in excess of 100% as macro economic crises (Sachs and Warner 1995), others defined periods of low or negative growth as macro economic crises (Lustig 2000), and again others have used extraordinary fluctuations of growth rates to reflect a macro economic crisis (Bannister and Thugge 2001). Since macro economic crises, regardless of how they are defined, tend to follow currency or banking crises, the rest of this chapter focuses on the causes of and remedies for these crises.

### **Causes of Banking and Currency Crises**

The standard approach to financial market development proposed by the international financial institutions, such as World Bank or IMF, is financial liberalization. The

underlying theoretical argument is that financial market deregulation improves an economy's efficiency, by eliminating financial repression. The concept of financial repression originally referred to the argument that government rules and regulations reduced the efficiency of the domestic financial sector in emerging economies, as not enough money was saved and as investors thus did not get enough funding for their investments (McKinnon 1973; Shaw 1973). More international capital mobility and less domestic regulation should hence improve the real economic performance of an economy, by allowing the financial sector to collect more funds and to distribute them more efficiently.

In contrast, though, some economists have argued that especially external liberalization increases competitive pressures on domestic banks, and induces them to accept greater portfolio risks than they would in absence of international competition (Demirgüç-Kunt & Detragiache 1998). Similarly, some sceptics of financial liberalization have argued that greater inside and outside liberalization increases the chance for financial crises as it entices investors to direct funds into speculative projects often driven by "deregulation euphoria" (Gabel 1993; Arestis & Demetriades 1999; Weller 2001).

Financial liberalization is generally understood as the deregulation of external capital flows and of domestic financial markets. External liberalization includes the reduction or elimination of controls on capital flows into and out of a country, which raises the possibility for more short-term capital flows and more foreign direct investment (FDI) into and out of an economy. In comparison, the liberalization of domestic financial markets include the



elimination of credit ceilings, lending requirements, and entry restrictions, and the widening of the operational scope of financial market participants. For instance, banks may be allowed to sell insurance, and investment bankers may be allowed to provide commercial loans (Litan et al. 2001).

The liberalization of financial markets is usually proposed in response to an actual or perceived lack of financial capital. The goal is to eliminate financial constraints for businesses, so that more resources become available for investment in plant and equipment. Banks are expected to become more efficient following deregulation, which should make more credit available for investments. In particular, banks should find it easier to attract deposits after the elimination of interest rate ceilings and to borrow capital overseas once controls on capital inflows have been lifted. Also, banks should be more inclined to lend without interest rate restrictions and with greater financial market competition. The liberalised allocation of financial resources supposedly directs funds towards their most efficient uses, thus boosting investment, productivity and growth.

The theory of financial liberalization, though, misses important dimensions of the way the world works. For instance, it has been recognised that the level of capital a bank has can influence its lending behavior and its stability (Stiglitz 1994). Banks with low capital and lower than expected earnings may seek out high risk, high return projects, thereby becoming unstable. A widely noticed example was the U.S. savings and loan (S&L) crisis in the late 1980s, when poorly capitalised banks undertook high risk real estate ventures. In emerging economies, greater financial market competition,

particularly from well-capitalised foreign banks may set this process in motion by lowering the profitability of domestic banks. In fact, “an increase in the share of foreign banks leads to a lower profitability of domestic banks” (Claessens et al. 1998). Poorer profit expectations may even lower a bank’s franchise value as to leave it *de facto* bankrupt. Such banks stand to lose little or nothing from taking on greater risks and by engaging in speculative investments, mainly in the stock market or the real estate market.

More intense competition is not the only factor that causes greater instability after liberalization. Financial firms may also find more investment opportunities, and they may find themselves driven to more speculative investments either through more intensive competition or in an atmosphere of “deregulation euphoria” (Arestis & Demetriades 1999).

A crucial problem arises because opportunities to engage in more speculative financing are likely to increase after financial liberalization because of financial deregulation, which follows a period of real improvements, but which also gives rise to speculative bubbles. In particular, based on Minsky’s (1986) “financial instability hypothesis”, some economists have argued that greater liberalization is likely to result in more speculative investments, and in more high risk, high return investments, with destabilising consequences for the entire economy (Grabel 1993; Weller 2001).

In this view, financial deregulation leads to short-term economic gains, and hence fuels optimistic expectations, and thus speculative activities. After financial deregulation, liquidity improves and more funds are available for productive and speculative purposes. The increase in investment opportunities is generally

matched by a decline in financial market regulations. An expanding real sector flush with capital, booming asset markets, increasing rates of return point towards an improving economy. With higher real interest rates and with expanding real and financial sectors, more funds are then attracted from overseas. More capital inflows, lead to a real currency appreciation, hence attracting even more capital.

Unfortunately, changes in economic fundamentals after liberalization merely improve the economic situation in the short-run. A continued currency appreciation helps to attract capital, which follows the promise of short-term gains in deregulated financial markets. Both a continued overvaluation and the diversion of funds for speculative purposes, particularly in stock or the real estate markets, generate the illusion of a sound and improving economy. Thus, otherwise well-capitalised, and sound banks are tempted to extend credit beyond prudent limits. While the real sector is hurt by a currency appreciation, by deteriorating terms of trade, and by the lack of credit (since it is going to other—more speculative—uses), financial markets still expand liquidity for speculative purposes in asset markets. The stock market and the real estate market are doing well, while simultaneously real production slows.

More financial speculation raises the chance of financial crises because banks face a larger downside risk. In particular, as the real sector slows down, debt-to-equity ratios are likely to rise, and the chance of bankruptcy increases. Similarly, asset market speculation means that at some point speculators' optimistic profitability expectations are not met, and they are unable to meet their own financial obligations. With rising defaults, international investors

become more likely to withdraw their short-term funds, further weakening the economy. Rapid capital outflows translate into a lack of funds for ongoing projects, thus fuelling an economic downturn in all sectors of the economy. Capital outflows reverse the currency appreciation, thereby adding to the burden of those who owe loans denominated in foreign currency. With further increasing loan defaults, a downward spiral is set in motion that depresses financial and non-financial sectors alike.

It is important to understand that firms are in a precarious situation before a crisis occurs due to the emphasis on short-term returns that makes long-term financing harder to obtain. If investors can gain significant returns on speculative ventures in a short period, less funds will be allocated to more long-term productive investments, where investors have to be patient. Thus, firms will generally find it harder to secure long-term external financing after deregulation. Since especially smaller firms depend on external financing, the allocation of funds away from productive uses is likely to hurt small and medium-sized enterprises (SMEs) or start-ups more than large corporations. Consequently, productive investments may not be undertaken, and firms can lose their competitive advantages, thereby possibly increasing the default risk for banks, and fuelling the flames that ultimately lead to the firestorm of economic crises.

Adding to the vicious cycle is the fact that the government is caught in an unenviable situation. To avoid a financial crisis, governments often revert, at least temporarily, to raising interest rates to keep short-term capital from leaving. Thus, monetary authorities add to the growing burden for borrowers, raising in turn the risk

that borrowers will not be able to repay their loans. Similarly, once a crisis is set in motion governments face increased demands on their budgets that have already been strained because of a slowdown in the real sector.

The empirical evidence on the connection between financial liberalization and financial crises has been mounting. In a survey of banking sector problems the IMF found that two-thirds of its member countries have experienced banking sector problems between 1980 and 1996 i.e., the period when financial deregulation found widespread acceptance (Lindgren et al. 1996). Similarly, in a summary of recent studies on capital mobility, Blecker (1999) found that at least for emerging economies there is strong evidence that increased capital mobility raises the chance of crises. Further, Kaminsky and Reinhart (1999) found that financial liberalization often preceded a banking crisis. Similarly, Grabel (1998) showed that increased financial fragility was a systematic occurrence after financial liberalization for emerging economies following the Asian crisis. In an econometric study of 26 emerging economies, Weller (2001) found that countries are much more likely to experience banking and currency crises after financial liberalization than before because their economic structure has become more susceptible to financial market risks. Lastly, the IMF devoted part of its September 2002 World Economic Outlook (IMF 2002) to the issue of global integration and increased volatility, while accepting the premise that trade integration resulted in more macro economic volatility and financial integration to more volatile capital flows.

Even though financial liberalization may offer some advantages, most of them are

short-lived in emerging economies, which have not developed the domestic institutions to control swings in capital flows. For instance, liquidity constraints are likely to be reduced after liberalization. But since a non-trivial share of additional funds may find its way into speculative financing, not all businesses may benefit equally from the additional liquidity. Similarly, liberalization offers more access to new economies for multinational businesses. However, if a crisis occurs, demand in these economies may be depressed for lengthy periods. And the liberalized environment may offer some growth opportunities, but in a more volatile environment as exchange rates, credit supply, and interest rates are prone to fluctuate more after liberalization than before. Thus, if some form of liberalization is needed to arrive at more efficient financial systems, it has to be done under avoidance of the drawbacks.

### **Relying on Markets: Early Warning Systems**

Instead of pondering alternatives to liberalization, policymakers could consider institutional improvements that would allow for early detection of a looming crisis. If an effective early warning system could be developed, policies could be adjusted as needed.

A number of empirical studies on financial crises have set out to identify predictors of crises. In a seminal paper on currency crises, Eichengreen et al. (1995) found that currency crises across a sample of twenty OECD countries over the period from 1959 to 1993 exhibited strong regularities. In particular, they found that changes of monetary aggregates, budget deficits, foreign exchange reserves, exports, balance of payments deficits and inflation

had significant statistical predictive powers. Kaminsky et al. (1998) and Kaminsky and Reinhart (1999) applied a similar approach to five industrialized and fifteen developing countries for the period between 1970 and 1995. Both studies found that exports, the real exchange rate, the ratio of money supply to official reserves, and price indices were statistically significant in predicting currency crises. Further, Corsetti et al. (1998) and Kaminsky and Reinhart (1999) found that banking crises often preceded currency crises. Corsetti et al. (1998) proxied the stability of the banking sector by the bad loan ratio, whereas Kaminsky and Reinhart (1999) relied on similar macro economic indicators for banking crises as for currency crises. Finally, Eichengreen et al. (1996) found evidence of contagion effects as the likelihood of a currency crisis occurring increased when a crisis has occurred elsewhere.

But how valuable are the empirical studies in predicting crises? Berg and Patillo (1998) analyzed three studies, Kaminsky et al. (1998), Frankel and Rose (1996), and Sachs et al. (1996), to assess their potential predictive power using the example of the Asian currency crisis. Frankel and Rose (1996) based their study on annual data from the IMF for more than 100 developing countries, which had the shortcomings that their crisis indicator cannot account for speculative attacks and that annual data are likely to miss short-term developments. Sachs et al. (1996) studied macro economic variables in 20 countries during the Mexican peso crisis. Since it was only based on one crisis and its global fall-out, albeit in a very detailed fashion, it may not be a suitable basis for predictions of crises in other countries. Kaminsky et al. (1998) set out to find early warning signals for currency

crises. Their research reviewed 25 earlier studies on currency crises and identified statistically significant indicators for crises. Consequently, they selected fifteen indicators (out of a possible 103) based on theoretical considerations and data availability. These indicators were then used to find empirical regularities among twenty countries over the period from 1970 to 1995. Berg and Patillo (1998) found only the Kaminsky et al. (1998) approach to yield reasonable prediction results. However, in each case there remained a large possibility of missing a looming crisis.

There are two reasons why early warning signals are hard to find. First, empirical studies regarding financial crises appear to have limited success because adequate and timely data are often not available, which is especially apparent in the studies on banking crises. Second, even if empirical research were successful in finding adequate indicators for looming crises, it is not clear whether policymakers could use this information to avoid a crisis. Early warning signals may simply inform policymakers of the unavoidable. If policymakers decide to use a particular early warning model, all market participants are likely to be aware of that. Hence, market participants are likely to act on the information of early warning models provide, thereby leading to “self-fulfilling prophecies”.

### **Stabilizing Emerging Economies Through Local Financial Institutions**

Instead of focusing on developing early warning systems for financial and economic crises, researchers have begun to concentrate on identifying potentially stabilizing institutions. These institutions may include developing local financial institutions, or capital controls.

One possible alternative, or addition, to financial liberalization is the development of domestic financial systems. Such domestic developments could be substitutes for or complements to capital account liberalization. However, two issues need to be addressed. First, financial institutions should receive priority over capital markets. Second, stable financial markets require public support, either in the form of prudent supervision and regulation or in the form of government subsidies.

Capital markets can provide some funds for investments, but their most important role appears to be the transfer of ownership. The U.S. equity market, for instance, has not been a net source of funds, but a net drain on funds as net equity issues in the corporate sector were negative in every quarter except one between 1994 and 2001. Obviously, smaller companies and start-ups have no or little access to capital markets and hence have to rely on financial intermediaries, such as banks, even more so than larger, more well-established firms. Thus, most companies need to rely on banks and other external finance providers, such as venture capitalists, to a larger degree than on capital markets as a source of external funds for investment. The example of the advanced economies suggests that in developing economies, the development of local finance providers needs to take priority over the development of deep and broad capital markets (UN 1999).

Developing stable local financial institutions, though, depends on public support. Particularly when financial markets are deregulated, regulatory and supervisory institutions need to be strengthened because of the greater chance of destabilizing, or even fraudulent activities. Further, the development of local institutions, such as

credit unions, co-operative banks, savings banks, or postal savings unions, can provide more funds without increasing financial fragility. Postal savings unions, for instance, played an important role in Japanese development by channeling the funds from large numbers of small deposits to large-scale development projects. These institutions, which can be restricted from engaging in speculative activities, can be used to stabilize the economy. However, because these institutions tend to serve a large number of small clients, their operations are often costly. To be able to compete, especially if interest rates are deregulated, the higher costs that these institutions face from servicing a large number of relatively small clients require public subsidies. Subsidies can come in the form of office space for postal savings unions, or in the form of tax credits. Credit unions in the US, for instance, enjoy a tax-free status. Also, German savings banks are government guaranteed, which allows them to offer credit at below market rates.

### **Role of Capital Controls**

A country's ability to develop financial institutions and use public policies to stabilize its financial markets is limited by a lack of capital controls and by its participation in international agreements. For instance, in 1999, some aspects of the design of Germany's savings banking system were challenged under EU rules, pitting an international agreement against national development interests. Similar conflicts may arise under GATS. Also, government subsidies tend to be evaluated negatively by international capital markets. If a government wants to continue subsidizing local financial institutions, and if controls on external capital flows have been

reduced, a country's sovereign bond ratings may be lower than otherwise, resulting in higher interest rates, less investment and slower growth.

Financial market development hence becomes a balancing act between developing local financial institutions and the need to attract foreign capital. The public policy issue is then to determine which capital flows are desirable and which ones are not. The distinction between portfolio investment and foreign direct investment (FDI) is typically of considerable importance to such an analysis.

Portfolio investment provides capital to the bond and stock markets from abroad. Whether portfolio investment increases the availability of new capital to firms depends on a number of factors. If, for example, foreign investors merely buy existing shares on the stock market, firms will not necessarily receive more funds. Further, if foreign investors purchase bonds there is no mechanism that prohibits firms from using these funds for speculation. Short-term capital flows, or "hot money" are often found at the core of financial market volatility. Hence, policy makers need to possess the tools to slow down the flow of portfolio investment and to encourage more long-term capital flows.

It is recognized, especially in light of the Asian financial crisis, that a country should have the right to control capital flows if they become a danger to its economic stability. The unilateral imposition of strict capital controls by Malaysia in the fall of 1998 has served as a case study for the use of capital controls during currency crises (Ariyoshi et al. 2000).

Countries can impose a variety of capital controls. For one, countries can impose minimum stay requirements. International

investors would be prohibited from withdrawing their funds prior to a pre-set time limit; capital could only be withdrawn gradually, thereby helping to avoid a financial panic. Another approach are so-called "Tobin taxes", or international capital transactions taxes, which would levy a penalty on short-term capital withdrawal, while impacting longer term capital to a lesser degree. Chile's unremunerated reserve requirement (URR) that was in effect between June 1991 and September 1998 constituted an "asymmetric Tobin tax" that was levied only on capital inflows. The URR was designed to make international loans with maturities of less than 90 days more expensive than loans with a greater maturity. Third, countries could also impose outright prohibitions of certain types of capital movements. Countries could require that profits earned on FDI are reinvested in the host economy.

Evaluations of the effectiveness of capital controls are rare. In a collection of fourteen country studies, Ariyoshi et al. (2000) provided a preliminary evaluation of the effectiveness of capital controls in terms of stabilizing emerging economies. They found that capital controls cannot substitute for sound macroeconomic policies; that no single measure can always be effective everywhere; that targeted controls leave sufficient room to be circumvented, and hence are likely to be less effective than comprehensive controls; and that the choice of controls is determined by the administrative capacity of a country.

However, the research did not consider the potential trade-off between greater stability and slower growth following the lack of capital. Klein and Olivei (1999) found, based on IMF data, that developed countries with open capital accounts were

also more likely to grow faster from 1986 to 1995 than countries with closed capital accounts. Although their results did not find a link for developing economies, these findings may be subject to revisions as more and better data become available. In particular, Edison et al. (2002) concluded, based on a sample of 57 countries, that greater financial integration, which includes fewer capital controls, may be associated with faster growth.

Besides controlling short-term capital, policymakers have also focused on attracting FDI. Some FDI includes the physical relocation of technology. But how much of this technology will benefit the host economy depends on national regulations. Wholly owned subsidiaries, for instance, are likely to guard their technological advantages very closely, so as not to nurture competitors. In the case of joint ventures between foreign and local partners, such proprietary control is less likely. Public policies can encourage technology transfers, for example, by requiring that foreign investors partner with local businesses by setting limits on the share of a local business that foreign residents can own. Similarly, host economies can prohibit foreigners from owning real estate.

### **Conclusion and Further Research**

The theoretical and the empirical literature has linked financial market deregulation to crises, largely because the opportunities for speculative investments increase in a deregulated environment.

Financial and economic crises not only affect the respective economies, but also its trading partners through trade and capital linkages. For instance, businesses in trading partner countries lose export markets and gain import competition, as exemplified by

the widening of the US trade deficit in 1998 and 1999 following the Asian crisis.

Consequently, the policy focus has centered on financial stability, while still providing emerging economies with needed financial resources. A market based approach has been to identify early warning signals of looming crises. There has been only limited success in developing early warning indicators, particularly because of the lack of adequate, timely and comprehensive data.

Instead, alternatives to financial market deregulation have been considered. These alternatives include a variety of capital controls, possibly in combination with public subsidies for local financial institutions, such as credit unions or savings banks.

Aside from developing and strengthening local financial systems, either through subsidies or capital controls or both, other policy options could be considered. Recent theoretical and empirical developments suggest that workers rights can help to improve productivity growth as worker turn-over is lowered and employer commitment to employees is strengthened. At the same time, better worker rights may help to equalize the distribution of economic resources between labor and capital and within labor. Consequently, worker rights may help to avoid overcapacities and thus lower the chance for speculative investments.

Although initial findings support a positive relationship between worker rights and financial stability (Weller & Singleton 2002), additional research is necessary. In particular, it is not clear whether the rise in stability results from faster productivity growth or from more equitable income distributions or from both. Also, some of the

empirical research in this area indicates that greater external liberalization, and the resulting rise in capital mobility, may reduce the development or the use of worker rights. And lastly, the implementation of worker rights as an international public policy issue raises many important questions (Heintz 2002).

## Selected References

- Arestis, P., and P. Demetriades. (1999) "Financial Liberalization: The Experience of Developing Countries", *Eastern Economic Journal*, Volume 25, Number 4, pp. 441-457.
- Ariyoshi, A., K. Habermeier, B. Laurens, L. Otker-Robe, I. Canales-Kriljenko and A. Kirilenko. (2000) "Country Experiences with the Use and Liberalization of Capital Controls", *IMF Occasional Paper* Number 190. Washington DC: IMF.
- Bannister, G.J. and K. Thugge. (2001) "International Trade and Poverty Alleviation", *IMF Working Paper* Number 01/54, Washington, DC: IMF.
- Berg, A. and C. Pattillo. (1998) "Are Currency Crises Predictable? A Test", *IMF Working Paper* #154/98. Washington DC: IMF
- Blecker, R. (1999) *Taming Global Finance*. Washington DC: Economic Policy Institute.
- Claessens, S., A. Demirgüç-Kunt and H. Huizinga. (1998) "How Does Foreign Entry Affect the Domestic Banking Market?", *Policy Research Working Paper* #1918, Washington DC: World Bank.
- Corsetti, G., P. Pesenti and N. Roubini. (1998) *Fundamental Determinants of the Asian Crisis: A Preliminary Empirical Assessment*, mimeo, Yale University.
- Demirgüç-Kunt, A. and E. Detragiache. (1998) "Financial Liberalization and Financial Fragility", *Policy Research Working Paper* #1917. Washington DC: World Bank.
- Edison, H., R. Levine, L. Ricci, and T. Slok. (2002) "International Financial Integration and Economic Growth", *NBER Working Paper* Number 9164. Cambridge MA: National Bureau of Economic Research.
- Eichengreen, B., A. Rose and C. Wyplosz. (1995) "Exchange Market Mayhem: The Antecedents and Aftermath of Speculative Attacks", *Economic Policy*, October 1995, pp. 251-312.
- Frankel, J. and A. Rose. (1996) "Currency Crashes in Emerging Markets: An Empirical Treatment", *Journal of International Economics*, Volume 41, pp. 351-366.
- Grabel, I. (1993) "Fast Money, "Noisy Growth": A Noise-Led Theory of Development", *Review of Radical Political Economics*, Volume 25, Number 3, pp. 1-8.
- Heintz, J. (2002) *Global Standards, Worker's Rights and Living Wages*. Unpublished manuscript, Amherst, MA: Political Economy Research Institute.
- International Monetary Fund. (IMF) (2002) *World Economic Outlook*. Washington, D.C.: IMF.
- Kaminsky, G. and C. Reinhart. (1999) "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems", *American Economic Review*, Volume 89, Number 3, pp. 473-500.
- Kaminsky, G., S. Lizondo and C. Reinhart. (1998) "Leading Indicators of Currency Crises", *IMF Staff Papers*, Volume 45, Number 1.
- Klein, M. and G. Olivei. (1999) "Capital Account Liberalization, Financial Depth, and Economic Growth", *NBER Working*



- Paper* Number 7384. Cambridge MA: National Bureau of Economic Research.
- Lindgren, C., G. Garcia and M. Saal. (1996) *Bank Soundness and Macroeconomic Policy*. Washington DC: International Monetary Fund.
- Litan, R., P. Masson and M. Pomerleano. (2001) *Open Doors: Foreign Participation in Financial Systems in Developing Countries*. Washington, DC: Brookings Institution.
- McKinnon, R. (1973) *Money and Capital in Economic Development*. Washington DC: The Brookings Institution
- Minsky, H. (1986) *Stabilizing an Unstable Economy, A Twentieth Century Fund Report*. New Haven & London: Yale University Press.
- Sachs, J., A. Tornell and A. Velasco. (1996) "Financial Crises in Emerging Markets: The Lessons from 1995", *Brookings Papers on Economic Activity*, Number 1, pp. 147-215.
- Sachs, J. and A. Warner. (1995) "Economic Reform and the Process of Global Integration", *Brookings Papers on Economic Activity*, Number 1, pp. 1-95.
- Shaw, E. (1973) *Financial Deepening in Economic Development*. New York: Oxford University Press.
- Stiglitz, J. (1994) "The Role of the State in Financial Markets", in Bruno, M.; Pleskovic, B. (Editors), *Proceedings of the World Bank Annual Conference on Development Economics. Supplement to The World Bank Economic Review and The World Bank Research Observer*. Washington, DC: World Bank.
- U.N. (1999) *World Economic and Social Survey*. New York NY: United Nations.
- Weller, C. (2001) "Financial Crises after Financial Liberalization: Exceptional Circumstances or Structural Weakness?", *Journal of Development Studies*, Volume 38, Number 1, pp. 98-127.
- Weller, C., and L. Singleton., 2002, "Political Freedom, External Liberalization, and Financial Stability", *EPI Technical Paper* Number 262. Washington, DC: EPI.

Christian E. Weller,  
 Centre for American Progress  
 Washington, DC, USA  
 cweller@americanprogress.org

## Economic Growth

Amitava Krishna Dutt

### **Introduction**

Economic growth is generally measured by the rate of growth of per capita real income or production (for instance Gross Domestic Product or GDP). By definition, higher growth makes countries achieve higher levels of per capita income in the future. For instance, a rate of growth of 1 per cent a year makes per capita GDP double every 70 years, while a growth rate of 5 per cent does this in less than 17 years. Since the main goal of economists and economic policy makers is to improve economic well being, of which per capita income is thought of as a major indicator and determinant, economic growth has long been a central issue in economics.

The origin of modern economic growth is generally dated back the second half of the eighteenth century in the UK with the Industrial Revolution (although long periods of slower growth were experienced earlier in various parts of the world). Growth then spread to other countries in continental Western Europe, and subsequently to the US, other countries of recent settlement, Russia and Japan. In the period from 1950 to 1973, sometimes referred to as the Golden Age, most developed countries grew at rates higher than previously, but after 1973 their growth slowed down. Many less-developed countries (LDCs) also experienced respectable growth during the Golden Age. The East Asian Newly-Industrialized Countries (NICs), that is, South Korea, Taiwan, Singapore and Hong Kong, grew at spectacularly high rates especially after the 1960. From the 1980s several Asian countries, including India, and especially

China, are experiencing high rates of growth. But for many Latin American and Africa countries the growth performance is poor, with several African countries experiencing negative growth.

### **The Beginnings of Growth Analysis**

Economists have tried to analyze the growth process from almost when modern economic growth began. The classical economists focused on the process of capital accumulation due to saving out of profits: given a technologically-given capital-output ratio the growth of capital resulted in the growth of output. Adam Smith (1776), who pioneered this approach, was optimistic about the continuation of growth because the increasing division of labor made possible by the expansion of the market and output would counter any negative effects on profits from rising wages caused by employment growth. He favored *laissez faire* and the removal of artificial barriers to markets, imposed by governments, which could impede growth.

David Ricardo (1817) also stressed the primacy of saving (which was automatically invested) and capital accumulation in the growth process, arguing that labor shortages would be removed by population growth responding to higher wages. However, he predicted that the economy would be led to a stationary state due to the extension of agricultural production to increasingly less fertile lands, which increased rent going to landlords (who did not save) and thereby squeezed profits. He recommended postponement of the stationary state through cheap food imports and hence advocated free trade. Thomas R. Malthus (1820) departed from the earlier classical notion that all saving is automatically invested, a proposition which has come to be called

Say's law, arguing that general gluts of overproduction were possible, and believed that increasing consumption spending by landlords and others promoted growth by increasing the demand for goods.

Karl Marx (1867), while following the tradition of emphasizing the roles of saving and capital formation, focused on the antagonistic relation between workers and capitalists, the importance of land and rent declining as the economy expanded. For Marx, class struggle between workers and capitalist determines the real wage of workers, and the existence of a reserve army of the unemployed makes the bargaining power of workers and their wages low (see Marglin, 1984). The surplus above wages (and material costs) accrues as profits to capitalists who save and invest, so that—assuming away the depreciation of capital for simplicity—the growth of capital and output is given by

$$g = s_c (1-wa)/v,$$

where  $s_c$  is the saving rate of capitalists out of profits,  $w$  the real wage and  $a$  and  $v$  the fixed labor-output and capital-output ratios. In addition to stressing the importance of saving, the equation shows that growth is speeded up by technological change, shown by reductions in  $a$  and  $v$ , which increases profits. Marx also recognized that a number of factors could negatively affect growth: labor shortages could emerge, strengthening the bargaining position of workers, raising wages,  $w$ , and squeezing profits; capitalists could adopt labor saving innovations which would increase capital intensity of production,  $v$ ; and there could be overproduction, leading to a realization crisis, as discussed by Malthus and later by Keynesian economists.

The emergence of the marginalist school—represented by Walras and

Jevons—in the 1870s shifted the attention of economists from economic growth to issues concerning the allocation of given resources among alternative uses. However, interest in growth continued among a few, including Joseph Schumpeter, who stressed the role of the innovative entrepreneur in introducing new technology and bringing about growth, and the followers of Marx.

### Early Growth Theory

With high levels of unemployment during the Great Depression, there was a return from the microeconomics of resource allocation to macroeconomics, especially in the work of John Maynard Keynes (1936). But while Keynes was mainly interested in the short-run issue of unemployment equilibrium, his followers, including Harrod (1939), attempted to extend his theory of effective demand to the long-run analysis of growth, thereby giving birth to modern growth theory. Following Keynes, Harrod distinguished between the saving plans of households and investment plans of firms. He assumed that saving is a fraction of lagged income, investment is proportional to changes in income or output, that is, by the accelerator mechanism, and that the economy typically has unemployed labor. Harrod's model has three major implications. First, the economy's "warranted" rate of growth, at which planned saving is equal to planned investment, so that firms have no incentive to change their production and investment plans, is given by  $s/v$ , where  $s$  is the saving rate and  $v$  the capital-output ratio implicit in the investment function. Domar's (1946) closely-related model interpreted  $v$  as the technical capital-output ratio, and showed that  $s/v$  is the maximal rate of growth of a capital-constrained economy, so that growth

could be increased by increasing the saving rate or by reducing the capital-output ratio (by improving the efficiency of capital use, through technological change, or by adopting more labor intensive techniques). Second, the economy was poised on a knife-edge around this growth rate: if the actual rate of growth of the economy was lower (higher) than it, investment would be less (more) than saving, firms would reduce (increase) the rate of change of output, taking the economy further away from the warranted rate. Third, assuming a fixed labor-output ratio,  $a$ , and a fixed rate of growth of labor supply,  $n$ , there was no guarantee that the economy, even if it grew at its warranted rate, would be able to grow at its natural rate given by the rate of growth of labor supply: if  $s/v$  is less (greater) than  $n$ , labor demand grows slower (faster) than labor supply, increasing (reducing) the unemployment rate. This problem has been called Harrod's long-run problem. Harrod's growth theory is therefore pessimistic about the ability of the capitalist economy to experience stable, full employment growth without government intervention, which is what Harrod recommended.

Growth theories immediately following his modified some of his assumptions and obtained less pessimistic conclusions. The neoclassical theory of growth of Solow (1956) and Swan (1956) solves Harrod's long-run problem by assuming that capital-labor substitution (which changes  $v$ ) by cost-minimizing, perfectly competitive, firms (departing from Harrod's assumption of a fixed capital-output ratio) maintains full employment during the growth process. However, as explicitly noted by Solow, it assumes that saving and investment are identically equal, thus side-stepping the knife-edge instability problem. Solow's

model assumes a neoclassical production function with smooth substitution between capital and labor, constant returns to scale (doubling all inputs doubles output), and diminishing returns to labor and capital. A constant fraction,  $s$ , of income is saved, all markets operate smoothly, and labor supply grows at a constant rate,  $n$ .

At a point in time, given stocks of capital and labor, all markets clear through price changes, so that labor and capital are fully employed, implying a given level of production and income. The dynamics of the economy can be analyzed in Figure 1 by examining how the capital-labor ratio,  $k$ , changes over time. Its growth rate,  $\hat{k}$ , is given by the difference between the growth rate of capital,  $\bar{k}$ , and the growth rate of labor,  $\bar{l} = n$ , which implies that

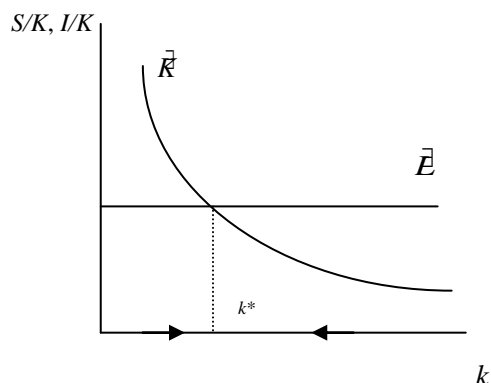
$$\hat{k} = s(Y/K) - n,$$

where  $Y$  is output and  $K$  is the stock of capital (the first term following from the fact that the rate of growth of capital is determined by saving – which equals investment – as a ratio of capital stock). Given the assumptions regarding technology, the output-capital ratio can be seen to depend on the capital-labor ratio: increases in capital, given labor, result in a reduction in the productivity of capital, reducing  $Y/K$ , so that the  $\bar{k} = sY/K$  is a downward-sloping line. The  $\bar{l}$  curve, given by  $n$ , is horizontal. It is assumed that the  $\bar{k}$  curve has the vertical and horizontal axes as asymptotes (which is ensured by the condition that both capital and labor are essential for production).

Given any initial value of  $k$ , the economy will then converge to the steady state level of  $k$ ,  $k^*$ , where  $sY/K = n$ , so that  $s/v = n$  is satisfied. At steady state,  $y$  becomes a constant as  $k$  is a constant, so that output per

worker or (output per capita, assuming for simplicity that everyone is a worker) becomes a constant, so that growth ceases. This occurs, however, because the model has abstracted from technological change. If it is assumed that technological change occurs at a constant rate, and that it is labor augmenting in the sense that in the production function it can be thought of as increasing effective units of labor,  $L$ , then we can simply reinterpret  $L$  as labor in efficiency units,  $k$  as capital per unit of effective labor and  $y$  as output per unit of effective worker. Then,  $n$ , the growth rate of effective labor supply is the sum of the rate of growth of labor supply and the rate of growth of labor productivity. Now, with  $y$ , output per effective worker becoming constant, the rate of growth of output per worker (and per capita) is given by the exogenously-given rate of growth of labor productivity.

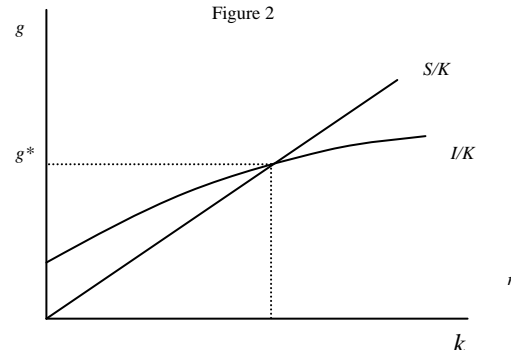
Figure 1



While the neoclassical model endogenizes the capital-output ratio,  $v$ , the Cambridge theory developed by Kaldor (1957) and others endogenize  $s$  by making the propensities to save from wages and profits to be different, as in Marx. Assuming for simplicity that workers do not save, the aggregate saving rate is given by  $s = s_c A/Y$ , where  $A$  is total profits and  $s_c$ , as before, denotes saving out of profits. The Cambridge model, like the neoclassical

model, assumes full employment growth, although Kaldor justified this by the supposed buoyancy of investment taking the economy to full employment, rather than neoclassical wage-price flexibility. If investment exceeds saving, there is an excess demand for goods, so that the price of goods increases, and if the money wage is constant or at least does not rise as much as the price, the profit share increases, increasing the saving rate,  $s$ , thereby bringing saving and investment to equality. Investment, determined in the long run by labor supply growth (due to the full employment condition), therefore determines saving in the model through changes in income distribution. Kaldor pointed out, however, that if income distribution cannot be changed due to wage or profit constraints, the adjustment mechanism need not work and unemployment could result.

Figure 2



The neoclassical and Cambridge growth models both assume that the economy grows with full employment at the rate determined by labor supply growth, differing only with regard to whether changes in the production technique or the saving rate allowed the economy to grow in such a manner. A third set of models—such as that of Robinson (1962)—follow the more Keynesian route of accepting the Harrodian conclusion about long-run growth and examines actual growth

paths which may not make the economy grow at its natural rate with full employment. Robinson (1962) assumed that desired investment depends positively on the rate of profit, that saving is a fraction of profits, and that growth equilibrium (in which saving and investment are brought to equality) is brought about by variations in the rate of profit as in the Cambridge model. The model is shown in Figure 2, where  $I/K$  is the desired investment curve, and  $S/K$  is the saving curve given by  $S/K = s_c r$  where  $r$  is the rate of profit,  $A/K$ . In this formulation, the knife edge instability problem is overcome by making saving respond more strongly to changes in the profit rate than investment (so that the investment curve is flatter than the saving curve). In growth equilibrium, capital grows at the rate  $g^*$ , so that, with a constant capital-output ratio and a given productivity of labor, output and employment also grow at rate  $g^*$ . If labor supply grows at rate  $n$ , there is no reason why this  $g^*$  should equal to  $n$ . Thus, unemployment may rise or fall over time. In this model growth is not determined by labor supply, but instead depends on business psychology. If firms' animal spirits are excited, the desired investment curve moves upwards and the rate of growth increases, with the profit share rising to allow saving to increase to meet the higher level of investment demand. Moreover, a rise in  $s_c$  shifts up the saving curve and reduces growth by reducing aggregate demand, the price and the profit share.

### **Empirics of Economic Growth**

The empirical analysis of economic growth, pioneered by Solow (1957), initially used time-series data from individual countries to conduct growth accounting exercises which decomposed growth into the growth of

inputs like capital and labor and technological change, usually measured as a residual. The subsequent literature has made use of the increased availability of comparable cross-country data to examine growth patterns across countries, looking both at growth differentials across countries, and at the determinants of cross-country differences in growth rates. The main results from these empirical studies can be summarized as follows (see Temple 1999; Durlauf et al 2005).

First, a large proportion of economic growth cannot be attributed to the accumulation of factor inputs such as capital and human capital (measured, for instance, by secondary school enrollment), and appears to be caused by technological change. However, since technological change and the increase in factors such as capital are likely to be interdependent, it is empirically difficult to disentangle these effects.

Second, on differences in growth performance across countries, much effort has been expended on whether poor countries grow faster than rich countries and therefore can catch up to them. These investigations, using a variety of techniques, have yielded little evidence of catching up. For instance, inequality across countries has been growing since the 1960s, and equations regressing subsequent growth rate to per capita income levels in 1960 show that on average poorer countries grew less rapidly than richer ones. There is evidence of convergence within the richer countries, but there is little evidence of many poorer countries joining this convergence club.

Third, regression methods have been used to explore the main determinants of economic growth using cross-section data. The number of proposed determinants in

these equations exceed 145, and include saving and investment rates, human capital, some measure of trade openness, export growth, government expenditure, military expenditure, foreign investment, income distribution, initial levels of per capita product, climate, distance from the equator, institutions and political instability. Average values of variables over a number of years have usually been considered in these equations to rule out short-run factors and concentrate on long-run influences, and in more careful studies, the possibility that growth can have a feedback effect on some of these variables has been taken into account in number of ways. Among the most robust results is the one that shows that saving or investment rates (as a ratio of GDP) have a positive effect on the rate of growth.

### **Neoclassical and New Growth Theory**

The neoclassical one became the dominant approach for the analysis of growth. The pioneering model of this tradition, the Solow model discussed earlier, has been extended in a number of directions (see Barro & Sala-i-Martin 1995). Solow's assumption that savers save a constant fraction of income has been replaced by making them into optimizing agents, either living for ever and maximizing utility over an infinite horizon, or maximizing utility over their two-period lifetime, as in overlapping generation models. These models have different implications for the stability of steady state growth and the efficiency (or Pareto optimality) of the growth path. While the infinite horizon optimizing model implies that the full employment growth path is stable and efficient, the arguably more realistic overlapping generations models can have different outcomes.

Despite its dominance, the neoclassical model seemed to be at variance with the empirics of growth discussed earlier. First, contrary to the observation that the rate of growth of per capita income depends positively on the saving and investment rate of the economy, almost all neoclassical growth models imply that the steady state rate of growth of per capita output is independent of factors such as saving and investment behavior, and dependent only on rate of technological change. This property can be demonstrated for the Solow model of Figure 1, where a rise in the saving rate makes the  $\bar{k}$  curve shift upwards, and increases the steady state levels of  $k$  and  $y$ , but not the rate of growth of the economy. The rise in the saving rate increases the rate of capital accumulation temporarily, but as diminishing returns to capital sets in, in steady state capital accumulation continues to be determined by the rate of growth of effective labor supply. Second, the model seems to suggest that richer countries will grow more slowly than poorer countries because of the greater force of diminishing returns in them. This, however, is not a necessary implication of these models, because higher saving rates for richer countries can make them grow faster even if they are closer to their steady states. Third, neoclassical models appeared to focus more on factor accumulation rather than on technological change, which is often taken to be exogenous, but which plays an important role in growth empirics. This criticism is not entirely fair, because technical change has been endogenized in a variety of ways in neoclassical models, for instance, by allowing productivity to increase due to learning by doing (see Arrow, 1962), research and development activity, and education expenditures.

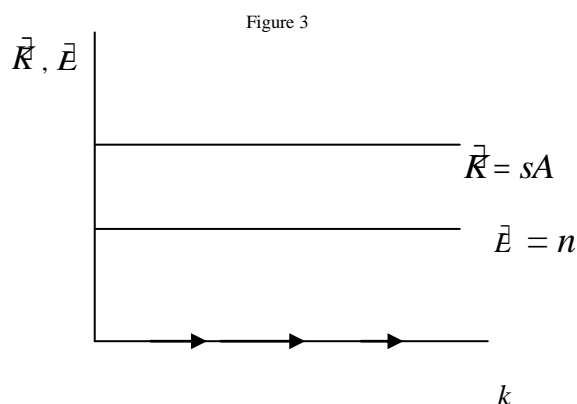
These weaknesses of the neoclassical model with diminishing returns led to the development of new or endogenous growth theories by Romer (1986), Lucas (1987), and others. This approach, which continues to follow the neoclassical tradition of assuming full employment, has now come to dominate growth theory. The essence of the approach can be shown with the simple AK model with the production function

$$Y = AK,$$

where  $A$  is the productivity of capital. This production function states that output increases proportionately with capital, without exhibiting diminishing returns, and is not affected by the amount of labor employed. Continuing with all of the other assumptions of the Solow model, we obtain

$$\dot{K} = sA - n.$$

Assuming that  $sA > n$ , the dynamics of the economy are shown in Figure 3. Starting from any initial value of  $k$ , the economy grows at the rate given by  $\dot{y} = \dot{K} = sA - n$ . An increase



in the rate of saving (and investment),  $s$ , implies a permanent increase in the rate of growth of the economy. Since capital, the produced means of production, is not subject to diminishing returns, an increase in the rate of capital accumulation raises the rate of growth of the economy in the long run. In fact, the absence of diminishing returns to

the produced means of production is not necessary for this result; all that is necessary is that there is a lower bound to the growth rate of capital which exceeds the rate of growth of labor supply.

Several interpretations to the AK production function can be given, following key contributions to the new growth literature. Romer (1986) takes investment to not only increase capital in the usual sense of private capital goods, but also the stock of knowledge, which is a public good and which increases the efficiency factor of labor, the two effects together implying non-diminishing returns to capital. Lucas (1987) can be interpreted as referring broadly to investment so as to include both physical and human capital; the growth of human capital increases the efficiency of labor, implying non-diminishing returns to capital. Yet other models countered diminishing returns to produced means of production by allowing increasing product variety and increasing returns within firms which produce these products (as distinct from externalities and public goods). The models with this form of increasing returns have to introduce imperfect competition to limit the size of firms, and this market power allows the consideration of innovative activity driven by profit-maximizing producers of knowledge who obtain temporary profits from new products (see Aghion & Howitt 1998). New innovations, in turn, increase the stock of knowledge on which future innovators can draw, but make obsolete products developed by earlier innovators.

New growth theory has important implications for many aspects of the growth process, of which we briefly discuss two. Given its focus on technological change, it is not surprising it has many interesting insights concerning it. A common



implication is that given the public goods nature of technological change (see Romer 1986), there will be underinvestment in the creation of knowledge in the free market economy compared to what is socially optimal; the implication is that government intervention can raise the rate of growth by speeding up technological change. A few models, however, suggest that it is also possible for the free-enterprise economy to overinvest in research and development, because the private research firm does not take into account its business stealing effect, that is, the fact it takes business away from firms which profited from an invention made obsolete by its invention (see Aghion & Howitt 1998).

It also has interesting implications for the effects of income distribution on growth. It may be supposed that since saving positively affects economic growth in these models, faster growth would result from increasing inequality, since the rich do most of the saving. If there are investment indivisibilities – for instance, large set-up costs for new industries – and if there are credit market imperfections, concentration of wealth among a few can, indeed, increase investment and the rate of growth. However, a number of models suggest that a more equal income distribution can increase growth (see Aghion & Howitt 1998). For instance, if there are no credit markets (or if these markets are imperfect), investors with higher endowments of wealth will invest more than those with less. If there are diminishing returns to capital, a misallocation of capital and a low level of aggregate output will result, which can result in a lower rate of technological change and growth (if aggregate output affects technological change due to learning by doing). Other models, such as those which

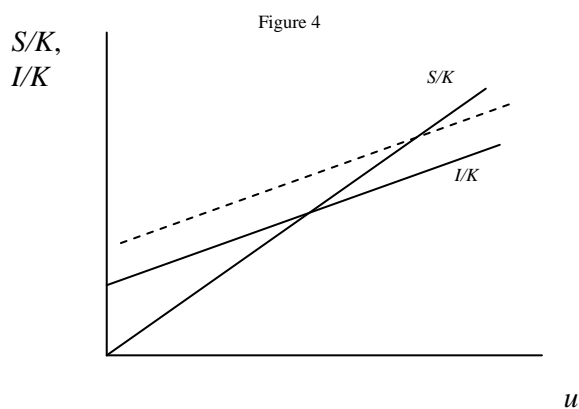
consider cooperation between the rich and poor and in which inequality leads to less cooperation, shirking and free riding by the poor, or in which inequality leads median voters to push for growth-reducing redistribution, produce similar results.

Both (old) neoclassical and new growth approach assume that all saving is automatically invested and that full employment always prevails. Even if unemployment exists in the short run due to wage rigidity, it is implicitly assumed that over the longer run, wage-price adjustments, by affecting the supply of money in real terms, or government stabilization policy, can take the economy to full employment, or at least to the “natural” rate of unemployment. Thus, aggregate demand can be ignored for the analysis of economic growth. However, it is increasingly being realized that wage-price adjustments may not take the economy to the “natural” level of output (because price deflation may reduce spending by borrowers and firms and because wage reduction can reduce consumption spending), the government may be unable or unwilling to do the job, and even if the economy does converge to the natural level of output, this level itself may be affected by aggregate demand and actual output (because unemployed workers may lose their skills and become incapable of effectively competing for jobs). If these arguments have some validity (see Dutt & Ros 2007), aggregate demand may well affect economic growth.

### **Aggregate Demand and Economic Growth**

Despite the dominance of new growth theory, some growth models have followed Harrod and, especially, Robinson in stressing the role of aggregate demand.

Perhaps the simplest model of aggregate-demand determined growth (see Taylor 2004 for more elaborate models) assumes that saving is a fraction of income, and that planned investment (as a ratio of capital stock) depends positively on the rate of capacity utilization, measured by the actual output-capital ratio,  $u=Y/K$ . Firms, who normally operate with excess capacity, increase investment when capacity utilization increases, both because it implies both higher profits and more robust markets. Output and, hence, capacity utilization adjust to bring aggregate demand and output, or saving and planned investment, to equality. The model, which is shown in Figure 3 with a linear investment function (shown by the solid investment line), is very much like the Robinson model, but whereas saving and investment are brought into equality through variations in the profit rate and income distribution, in this model the adjusting variable is capacity utilization. With the investment-capital ratio being determined as shown in the diagram, the growth rates of capital, and hence output and employment, are determined.



The model implies that a change in business psychology that excites “animal spirits” shifts the investment curve up (to the dashed investment line), and thereby increasing the rate of growth of the economy (including per capita income, if population

and labor supply growth are exogenously given). If government expenditure and monetary issues are introduced into the model, fiscal or monetary expansion by the government also results in an increase in the rate of growth of the economy. A rise in the saving rate, however, implies a fall in the rate of growth of the economy, because higher saving implies lower consumption demand, lower capacity utilization, and hence lower investment and growth.

This simple model can be criticized on a number of grounds. First, it appears to contradict the empirical finding that higher saving (and investment) rates result in higher rate of growth. A modified version of the model, which assumes that saving depends negatively on the stock of capital (or wealth), and on profits rather than income (under the assumption that all wage income is consumed), implies that a higher saving and investment rate (as a ratio of income) is associated with a higher rate of growth, both being increased by an increase in autonomous investment and by a *reduction* in the propensity to save out of profits (because it increases consumption demand). Second, the model determines the rate of capacity utilization endogenously, having the allegedly dubious implication that actual capacity utilization can deviate from planned capacity utilization in equilibrium. However, this implication can be found acceptable if it is assumed that firms do not have unique level of planned capacity utilization, so that the equilibrium may be consistent with a range of capacity utilization rates, at least within a band. Moreover, actual and planned capacity utilization may both be endogenous and equal to each other if planned capacity utilization depends on strategic considerations (Dutt 1997). Third, if labor supply grows at an exogenous rate, the

model implies that the unemployment rate can rise or fall indefinitely in equilibrium. This problem, however, can be overcome by allowing labor supply to change endogenously as a result of labor shortages, and alternatively, because of endogenous changes in technology which are also driven by labor shortages (see Dutt 2006). Models which introduce endogenous technological change of this type can therefore allow aggregate demand shocks to have permanent, long-run, effects on the rate of economic growth.

Aggregate demand based growth models have a variety of interesting implications, of which—following our previous discussion on new growth models—we consider those regarding income distribution and technological change. Income distribution is examined in these models by assuming, as in Marxian and Cambridge models, that there are different propensities to save out of profit and wage income. Suppose, as in Robinson's model, that a fraction of profits is saved, but all wages are consumed. Then, a rise in the share of profits due to, for instance, a rise in the price-cost markup charged by firms, has the effect of making the distribution of income less equal and reducing the rate of growth of capital stock and output. This occurs because the shift in income distribution towards profits increases saving, thereby reducing consumption and aggregate demand, reducing capacity utilization and hence the rate of investment and growth (see Dutt 1990). Such models, however, need not necessarily have this implication. For instance, if investment depends both on capacity utilization and the profit markup (since both are likely to affect profit expectations), a rise in the markup can increase both the profit share and the rate of investment and growth because of the direct

effect of the markup on investment (Bhaduri & Marglin 1990).

Regarding technological change, mention has already been made of how endogenous technological change can provide a mechanism of how the rate of growth of labor effective labor supply can be brought in line with the rate of growth of labor demand in response to aggregate demand growth. The approach also stresses that an expansion in aggregate demand, by speeding up economic growth, can increase the rate of productivity growth due to learning by doing. On the consequences of technological change this approach implies a more ambiguous role to technological change than in the neoclassical and new growth theory approaches. For instance, it is possible that technological change can have the effect of reducing the rate of growth of labor demand, and thereby increasing unemployment, rather than of increasing growth, and higher unemployment may even have the effect of reducing the share of wages in income and reducing aggregate demand and growth. On the other hand, technological change can have the effect of increasing investment demand because firms install new machinery to take advantage of newer production methods, and also increasing consumption demand due to product innovation, thereby increasing aggregate demand and economic growth. Thus, it is possible for there to be a two-sided relation between economic growth and technological change.

### **Unequal Growth and International Issues**

It was noted earlier that empirical research shows that inequality among countries has been growing since the 1960s and that countries with low per capita income, on average, have been growing more slowly than rich countries. In fact, historical data

suggest that the divergence between rich and poor countries has been occurring for a considerably longer period of time (Maddison 1991, Pritchett 1997).

Such unequal growth can be explained by factors internal to individual economies. For instance, if rich countries save a higher fraction of their income, then new growth theory implies they will grow faster, or if rich country firms are more optimistic about the future, they may invest more, and models in which aggregate demand affects long-run growth imply that they will grow faster.

However, the possibility that international growth patterns can be explained by interactions between countries has also been extensively analyzed, and this analysis is of great importance in assessing the implications of closer interactions between countries through international trade, capital flows, and other aspects of globalization. It is often argued that greater openness is good for growth, since export expansion and foreign direct investment inflows foster technological change, while import protection leads to inefficiency. Moreover, the removal of restrictions on international capital flows is argued to facilitate the movement of capital from capital abundant rich countries to capital scarce poor countries in search of higher profits, and thereby speed up capital accumulation and growth in poor countries. But greater openness, by increasing imports, may reduce aggregate demand, growth and technological change. Moreover, increased imports may result in the contraction of production in technology-intensive sectors in which learning by doing and skill formation are more important (see Dutt, 1990), thereby reducing the poor country's technological capability. The transfer of foreign

technology may be able to reverse this process, but successful technology transfers require that the recipient country is not too far, in terms of technological capability, from rich countries.

### **Policy Implications**

Although this brief overview has been able to discuss only some of the major analytical and empirical issues concerning economic growth, it has shown that growth is determined by a variety of factors, the importance and relevance of which may depend on the context and economy finds itself during a period of time. Thus, for instance, growth may sometimes be determined by saving, as in the classical and new growth approaches, by technology which limits the effective supply of labor as in the neoclassical and new growth approaches, by investment incentives as in aggregate demand models, and by income distribution as in new growth and aggregate demand approaches. The policy implications of this analysis are therefore complex, more so if one brings in issues we have omitted. Some comments about these implications are provided in terms of three classic policy debates concerning growth.

Concerning the debate on the efficacy of free markets versus government intervention for promoting growth, the case for free markets is made by the classical and neoclassical approach with infinitely-lived optimizing agents in terms of the markets promoting the division of labor and allocating resources in an efficient manner. Other models, including the Marxian approach in mediating class conflict, new growth theory models featuring externalities (especially concerning technological change and education), and aggregate demand models emphasizing unemployment and

demand deficiency, provide arguments for government intervention. The actual experience of fast growing East Asian economies suggests that governments can allocate investment to sectors which have high technological spin-offs to the rest of the economy, while exposing firms to the discipline of the market. The ability and willingness of governments in successfully promoting growth, of course, depends on political economy issues which go beyond the concerns of growth theory.

If the government is to intervene to promote growth, should it rely on supply or demand-side policies? The focus of traditional growth theory has been on the supply side, recommending policies to increase private and government saving, and to speed up technological change by promoting investment in education and research and development. In economies in which growth is determined by aggregate demand—that is, where there are unemployed or underemployed resources or where technological change is responsive to demand pressures—these policies may be counterproductive, and more of a focus on the aggregate demand side is called for, by strengthening investment incentives, and by avoiding sharply contractionary stabilization policies in response to macroeconomic problems such as financial crises. In many situations, some policies, such as those that lead to income distributional improvements, can promote growth from the demand side as well as from the supply side.

On the issue of economic openness in the form of the promotion of trade and capital flows, although it is possible that more open policies which favor globalization can help growth in poor countries, neither theoretical analysis or empirical studies suggest that they are a panacea. While countries which

have achieved a minimum level of technological capability may be able to take advantage of foreign technology and export opportunities, the recent experience of most Latin American and Africa countries suggests that neo-liberal policies may hinder growth in poor countries.

To conclude, one should ask whether economic growth is really desirable. By and large economists have not reflected on this question, either relying on the idea that more (of goods and services) is better, or because growth improves the lot of the poor. It is by no means clear, however, that more is really better: cross country evidence suggests that beyond a certain level of per capita income further increases do not have a significant impact on measures of subjective well-being obtained from questionnaire surveys, and similar results are found from time series studies of advanced countries like the US and Japan (see Frey & Stutzer 2002).

It is also possible that economic growth can have negative influences on other things that may be considered to be of value, such as the environment, and indeed, such influences may have adverse long-term effects on economic growth itself. While economic growth may be vital for reducing poverty in poor countries, one can ask whether redistributive policies rather than growth-enhancing ones are more important for tackling poverty in more advanced countries. On the one hand the pursuit of growth and efficiency can lead public policy away from focusing on social policies which directly affect the poor. On the other hand, economic growth can lead to greater openness of opportunity, commitment to fairness, tolerance of diversity and promote democracy (see Friedman 2005); with people becoming less interested in protecting their current positions, it may be

easier to divert resources to the poor when economic growth is more rapid.

### Selected References

- Aghion, Philippe and Peter Howitt. (1998) *Endogenous Growth Theory*. Cambridge, Mass: MIT Press.
- Arrow, Kenneth K. (1962) "The Economic Implications of Learning by Doing", *Review of Economic Studies*, Volume 29, pp. 155-73.
- Barro, Robert J and X. Sala-I-Martin. (1995) *Economic Growth*. NY: McGraw-Hill, 1995 (2<sup>nd</sup> Ed., Cambridge, Mass: MIT Press, 1999)
- Bhaduri, Amit and Marglin, Stephen A. (1990) "Unemployment and the Real Wage: the Economic Basis of Contesting Political Ideologies, *Cambridge Journal of Economics*, Volume 14, Number 4, Pp. 375-93.
- Domar, Evsey D. (1946) "Capital Expansion, Rate of Growth, and Employment", *Econometrica*, Volume 14, 137-47.
- Durlauf, Steven N., P. Johnson and Jonathan Temple. (2005) "Growth Econometrics", in P. Aghion and S. Durlauf (Editors), *Handbook of Economic Growth*. Amsterdam: North-Holland, Forthcoming.
- Dutt, Amitava Krishna. (1990) *Growth, Distribution and Uneven Development*. Cambridge, UK: Cambridge University Press.
- Dutt, Amitava Krishna. (1997) "Equilibrium, Path Dependence and Hysteresis in Post-Keynesian Models", in P. Arestis and M. Sawyer (Editors), *Essays in Honour of G. C. Harcourt, Vol 2: Markets, Unemployment and Economic Policy*. NY & London: Routledge.
- Dutt, Amitava Krishna. (2006) "Aggregate Demand, Aggregate Supply and Economic Growth", *international Review of Applied Economics*, Volume 20, Number 3, pp. 319-36.
- Dutt, Amitava Krishna and Jaime Ros. (2007) "Aggregate Demand Shocks and Economic Growth", *Structural Change and Economic Dynamics*, Volume 18, pp. 75-99.
- Frey, Bruno S. and Stutzer, Alois. (2002) *Happiness and Economics*. Princeton: Princeton University Press.
- Friedman, Benjamin M. (2005) *The Moral Consequences of Economic Growth*. New York: Alfred A. Knopf.
- Harrod, Roy F. (1939) "An Essay in Dynamic theory", *Economic Journal*, Volume 49, March, pp. 14-33.
- Kaldor, Nicholas. (1957) "A Model of Economic Growth", *Economic Journal*, Volume 67, pp. 591-624.
- Keynes, John Maynard. (1936) *The General theory of Employment, interest and Money*. London: Macmillan.
- Lucas, Robert E. (1988) "On the Mechanics of Economic Development", *Journal of Monetary Economics*, Volume 22, pp. 3-42.
- Maddison, Angus. (1991) *Dynamic Forces in Capitalist Development*. Oxford: Oxford University Press.
- Malthus, Thomas R. (1820) *Principles of Political Economy*. NY: Augustus M. Kelley, 1988
- Marglin, Stephen A. (1984) *Growth, Distribution and Prices*. Harvard: Harvard University Press.
- Marx, Karl. (1867) *Capital: A Critique of Political Economy*, Volume 1. Edited by F. Engels. Moscow Foreign Languages Publishing House, 1957.

Amitava Krishna Dutt  
Department of Economics and Policy  
Studies, University of Notre Dame, Notre  
Dame, Indiana, USA  
adutt@nd.edu

- Pritchett, Lant. (1997) “Divergence, Big Time”, *Journal of Economic Perspectives*, Volume 11, Number 3, pp. 3–17.
- Ricardo, David. (1817) *On the Principles of Political Economy and Taxation*. Cambridge: Cambridge University Press, 1962.
- Robinson, Joan. (1962) *Essays in the theory of Economic Growth*. London: Macmillan.
- Romer, Paul M. (1986) “Increasing Returns and Long-Run Growth”, *Journal of Political Economy*, Volume 94, pp. 1102-37.
- Smith, Adam. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: J.M. Dent & Sons Ltd, 1947.
- Solow, Robert M. (1956) “A Contribution To the Theory of Economic Growth”, *Quarterly Journal of Economics*, Volume 70, pp. 65-94.
- Solow, Robert M. (1957), “Technical Change and the Aggregate Production Function,” *Review of Economics and Statistics*, Volume 39, pp. 312-320.
- Swan, Trevor W. (1956) “Economic Growth and Capital Accumulation”, *Economic Record*, Volume 32, pp. 334-61.
- Taylor, Lance (2004) *Reconstructing Macroeconomics. Structuralist Proposals and Critiques of the Mainstream*, Cambridge, Mass.: Harvard University Press.
- Temple, Jonathan (1999) “the New Growth Evidence”, *Journal of Economic Literature*, Volume 37, Number 1, pp. 112-56.

## Employer of Last Resort Schemes

*Fadhel Kaboub*

### **Introduction**

The idea of government as the employer of last resort (ELR) has been present in economic literature since the seventeenth century. The need for an ELR program became more urgent after the industrial revolution. Capitalist economies lack an inherent mechanism to create full employment. The Great Depression was the worst episode of the system's failure to deliver full employment. At that time, John Maynard Keynes was one of the few economists who challenged the conventional wisdom by arguing that capitalism, when left to its own devices, will not gravitate towards full employment, and suggested that government intervention was required in order to jump-start the economy and help achieve and maintain full employment.

Unemployment had a devastating socio-political and humanitarian impact during the Great Depression. The United States started to recover only when the government began spending massively to support the war effort. But even before WWII had ended, economists began debating whether or not capitalist economies will fall back into another depression. In his influential book, *Full Employment in a Free Society*, Sir William H. Beveridge called on the government to guarantee full employment, which for him "means having always more vacant jobs than unemployed [people], not slightly fewer jobs. It means that the jobs are at fair wages, of such a kind, and so located that the unemployed [people] can reasonably be expected to take them; it means, by consequence, that the normal lag between

loosing a job and finding another will be very short" (Beveridge 1945:18).

At the onset of capitalism, Sir William Petty recognized that unemployment was a serious problem that needed to be addressed by society. Unlike many English businessmen and thinkers of his time, Petty believed that the unemployed "ought neither to be starved, nor hanged, nor given away" (Petty 1662:30). However, Petty was not the humanitarian that we would like to think he was; he simply believed that the unemployment pool was an untapped source of enrichment for the nation, and that the unemployed could be publicly employed to build infrastructure; "at worst this would keep their minds to discipline and obedience, and their bodies to a patience of more profitable labours when need shall require it" (Petty 1662:31).

By the time David Ricardo published the third edition of his *Principles of Political Economy and Taxation* (1821), the capitalist mode of production has already become an integral part of the economy. Ricardo saw the need to add a new chapter, *On Machinery* (chapter 31), in which he acknowledged that "the substitution of machinery for human labour, is often very injurious to the interests of the class of labourers" and added that "the same cause which may increase the net revenue of a country, may at the same time render the population redundant, and deteriorate the condition of the labourer" (Ricardo 1821:283-4). Ricardo died two years later (1823) and therefore did not live to fully elaborate on his 'new' views about capitalism and unemployment; but another great thinker, Karl Marx, came to the same conclusion about the chronic nature of unemployment in capitalist economies. For Marx the industrial reserve army of the



unemployed is a vital part of capitalism allowing the system to expand during a boom and keeping profits high during a recession by holding wages down with the threat of being discharged into the reserve army.

For more than three centuries, neoclassical economists have considered that unemployment is only a transitory phenomenon, and have either denied or minimized the existence of involuntary unemployment. Even during periods of high unemployment, they argued that the only thing the government should do is keep its hands off of the market which would eventually clear the labor market; furthermore, they advocated government should reduce its spending and encourage downward wage flexibility. The unemployment of the Great Depression, however, proved to be disastrous at all levels. The government had to do something.

In the United States, after several attempts to promote market forces, and a disbelief in Keynes's recommendations for increased government spending to boost aggregate demand, policymakers came to realize that *laissez-faire* economics was not the solution to the problem at hand (25% unemployment in 1933) and that the government had to act as the employer of last resort. In 1933, President Roosevelt introduced the New Deal program along with a whole host of public employment agencies that were enacted between 1933 and 1936, including the Public Works Administration, the Civil Works Administration, the Works Progress Administration, the Civil Conservation Corps, the National Youth Administration, Rural Electrification Administration, and the Federal Emergency Relief Administration.

Despite the immediate success of these agencies, business opposition to the New Deal programs grew strong and Roosevelt quickly backed down and promised the nation that "a balanced budget [was] on the way"; and so was unemployment, which reached 14.3% in 1937, and then rose to 19% in 1938 when Roosevelt decided to slash government spending as a response to inflation fears emerging from the Federal Reserve Bank, which also doubled reserve requirements between 1936-1937. For these reasons, the New Deal was never meant to be a true ELR program since it did not provide an infinitely elastic demand for labor, but it did empirically show that the government can act as employer of last resort and provide decent jobs that do not compete with the private sector and that are socially, economically, and environmentally useful.

A new generation of economists began to think about ELR as a serious policy alternative rather than just an intellectual exercise. And without too much discussion when WWII came along, the deed was done. Full employment was practically maintained throughout the war period with unemployment as low as 1.2% in 1944, the lowest rate ever recorded in the history of the United States. Concerns over accelerating inflation were taken seriously when the Office of Price Administration (1942-1947) used the March 1942 prices as a ceiling for nearly 90 of retail food prices, as well as on residential rent. Rationing was also imposed on key commodities during the same period. Full employment and price stability were achieved but only during wartime with a considerable number of the male working-age adults being in army and the rest of the working-age population employed to support the war effort.

Therefore, the U.S. full employment experiment must be taken with caution given the specific circumstances, and the results should be evaluated accordingly, even though there is a lot that could be learned from the methods employed at the time.

### **John Pierson's *Economic Performance Insurance***

John H.G. Pierson (1906-2001) was a Yale-educated economist who held several prominent positions in the U.S. Department of Labor and helped draft the Employment Act of 1946. Frustrated by the miseries of the Great Depression and by the unrealistic economics textbook models, Pierson began working on what would become his lifetime "obsession"—full employment. The result was published in his first book *Full Employment* (1941) and was later refined and further elaborated in his 1964 book *Insuring Full Employment: A United States Policy for Domestic Prosperity and World Development*.

The basic premise of his Economic Performance Insurance (EPI) proposal was that the government should adopt a policy of *guaranteed* full employment. In other words, the government should "stand ready to step in as employer of last resort; or step out, when necessary—[as...] disemployer of first resort. The mechanism to permit that would be a nationwide reserve shelf of additional public services and public works" (Pierson 1980:53). But this is only one side of the equation according to Pierson; the other side is indeed a very crucial one in maintaining full employment in the long run: high levels of consumer demand. It sure can be achieved through the employer of last resort, but one cannot rule out the possibility of oversaving to take place in the economy which will cause demand to decrease and as a result

employment in the private sector to fall as well, thus leading to unemployment again, unless if the number of last-resort jobs is constantly increased to absorb the sluggishness of the system. To avoid this scenario, Pierson's EPI proposal emphasizes the importance of the government *guaranteeing* or *underwriting* the volume of consumer spending that is consistent with the full employment level of production. That could be done through adjustments of consumer taxes or transfer payments, negative income tax, reversible federal sales tax, or federal sales bonus at the retail level, which could be stamps convertible into cash—"income boosters" or I.B.s. (Pierson 1980:54).

With his EPI proposal, Pierson showed that it is important not only to guarantee *current* aggregate demand (through the reserve shelf of public works), but also *future* markets which would boost expectations and therefore increase investment. Hence the double confidence-building feature of the EPI system would guarantee that there are enough jobs to be had, and that the market will be held at an adequate level year after year. Furthermore, EPI would kill off inflationary expectations. By establishing floors under employment and consumer spending as well as ceilings over both, no inflationary spiral could take place. If need be, the government can scale back on consumer spending subsidies and/or call for putting some public works back on the reserve shelf until further notice.

Even though Pierson's EPI proposal was put forward as a policy for domestic prosperity, it also has its international benefits as well. When domestic full employment is guaranteed through EPI, there would be less pressure on domestic producers to worry about finding more or

new markets at the international level since domestic demand and jobs are already *guaranteed*; thus leaving more room for developing countries to expand their exports and move from aid to trade (Pierson 1964, 188).

### **John Philip Wernette's *Full Employment Standard***

Like Pierson, John Philip Wernette (1903-1988) was fearful that the post-war era will be marked by mass unemployment and another Great Depression. In 1945, he published *Financing Full Employment*, in which he laid out a long-term economic policy proposal to secure full employment. Wernette proposed to establish a "new fiscal-monetary system," which he called the "Full Employment Standard" (FES). Recognizing that the capitalist system is inherently unstable, Wernette advocated the establishment of a Federal Stabilization Board which would "take over the powers and duties of the Federal Reserve Board of Governors, and acquire some new ones" namely the "control of the total amount of money", and "the creation of new money to be turned over to the Federal government to finance budget deficits and/or to pay off the Federal debt" (Wernette 1945:30-33).

Wernette believed that the Federal government has the ultimate responsibility of ensuring the conditions of full employment. Like Pierson and other ELR advocates, Wernette's FES proposal is carefully designed for a dominantly market economy. "The entire purpose of the program [...] is to keep private enterprise alive by underwriting a big market for the goods and services which private business can produce" (Wernette 1945:10).

The FES proposal is financed mostly by printing new money. This is in fact

necessary in order to achieve and maintain prosperity and full employment. And according to Wernette, an expanding economy (with a growing population and a rapidly rising potential per capita production) can absorb immense amounts of new money without creating inflation (Wernette 1945:34).

The orthodox dichotomy between fiscal and monetary policy is absent from the FES proposal. Indeed, it is the meticulous coordination between the two that lies at the heart of FES. The Federal Stabilization Board would create monetary certificates and deposit them in the Federal Reserve banks for the account of the U.S. Treasury. The Treasury then draws checks to finance deficit spending (without borrowing) on public works projects that are carefully planned and selected not to be in competition with private enterprise.

Wernette developed a very elaborate argument explaining how money is injected into the system through government spending and is destroyed through taxation. "The function of Federal taxes is preventing inflation. The federal government literally does not have to collect taxes in order to get the money for its expenditures. Like any other sovereign government, our Federal government has the power of creating money" (Wernette 1945:70-71). Wernette thus had rediscovered the principle of functional finance which Abba Lerner had laid out in his seminal 1943 *Social Research* article, "Functional Finance and the Federal Debt."

### **Abba Lerner's Functional Finance Theory**

Abba Lerner's (1903-1982) most important contribution to the literature is the principal

of “functional finance” which opposes the orthodox view of “sound finance.”

“The central idea is that government fiscal policy, its spending and taxing, its borrowing and repayment of loans, its issue of new money and its withdrawal of money, shall all be undertaken with an eye only to the *results* of these actions on the economy and not to any established traditional doctrine about what is sound or unsound” (Lerner 1943:39; emphasis in original). From a functional finance perspective, “money is the creature of the state.” Money is created when the government spends, and is destroyed when the government levies taxes. The government does not need to “borrow” its own money from the public; rather it only “borrows” in order to withdraw excess money from the system and to give savers an alternative interest-bearing asset (bonds).

Similarly, the government does not need to tax its population in order to finance expenditures; rather the government needs the public to demand its currency to give it value. Hence, there can be no financial constraint on the monopoly-issuer of money (the state). A sovereign state can make anything generally acceptable and call it “money”, as long as the state “is willing to accept the proposed money in payments of taxes and other obligations to itself” (Lerner 1947:313). From this perspective, all the worries about the deficit and the national debt become meaningless when compared to their *function*: financing full employment. The employer of last resort is in fact the spender of last resort whose responsibility is to keep the rate of aggregate spending in the economy “neither greater nor less than that rate which at the current prices would buy all the goods [and services] that it is possible to produce.” If the rate of spending is too

high, inflation will develop; and if it’s too low, there will be unemployment (Lerner 1943:39).

This taxes-drive-money (TDM) approach has been one of the cornerstones of most ELR policy proposals. The TDM theory has been challenged by Louis-Philippe Rochon and Matias Vernengo (2003) who claim that the state can only establish the validity of money, but not its value, and that bank (credit) money takes precedence over state money. They argue that when the state is weak and cannot enforce tax collection, banks may still create money through loans to finance the hiring of workers to produce new output. Money will be accepted and used in the economy because workers will use it to buy the output and firms will accept it because they can use money to pay off the loans they took from banks. The value of money is then determined by the interaction between the newly produced output and the number of money units that firms pay out to the workers who produce it. Although this is certainly a valid scenario, it does not, however, contradict the TDM approach; it merely shows that the *authority* that controls the creation and destruction of money, gives it value, and imposes its general acceptability can be either the state, the banking system, religious authorities, the military, or even the mafia. The loss of state sovereignty (broadly defined) creates a vacuum and allows other authorities to emerge and control monetary processes (see Wray 2004).

### **The Swedish Full Employment Model**

The Swedish (corporatist) model showed that there is a way to achieve price stability without using unemployment as a disciplinary measure against labor. The model was developed after WWII by trade

union economists Rehn and Meidner who envisioned two essential elements that would characterize the Swedish economy for more than four decades: 1) highly centralized wage bargaining; and 2) active labor market policies. The model focused on the ‘socialization of investment’ and offered a practical alternative to welfarism by putting a strong emphasis on “the right to work” rather than “the right to income.”

The strong trade unions stressed non-inflationary full employment and wage differentials based on skills and training rather than profitability in a given industry. Thus, equitable income distribution was a basic tenet of the Swedish model. A wage restraint on behalf of labor unions was compensated with a system that taxes profits and uses the proceeds to finance capital accumulation under the workers’ control. Swedish employers took the initiative on centralized wage bargaining in the 1950s with the potential wage-restraint gains and the ‘peace obligation’ (no strikes) as an end-in-view (Marshall 1995:213). Thus, a fundamental feature of the Swedish corporatist model was the unity of interest, which was manifested through a non-centralized wage restraint and centralized wage bargaining.

The National Labour Market Board (AMS) has played a key role in implementing activist labor market policies in Sweden. It is a tripartite institution (labor majority, business, and government) funded by Parliament via the annual budget appropriation. AMS meets twice a month to make decisions about labor exchanges, training programs, and wage subsidies for workers who have not been placed within six months. Only after all employment and training options have been exhausted are individuals entitled to unemployment

benefits, and once those benefits are also exhausted, the unemployed has the right by law to work up to six month in the public service employment which acts as the employer of last resort (Ginsburg 1983:127-154).

The theoretical underpinning of the Swedish model can be found in Keynes’s idea of ‘socialization of investment’. The model strongly encouraged private investment despite high tax rates on profits. Firms were allowed to put their ‘excess profits’ into tax-exempt ‘investment funds’ thus encouraging capital accumulation. These investment funds in fact date back to 1938 and have been used primarily as counter-cyclical tools rather than a true socialization of investment. In 1983, five independent regional wage-earner funds were created with majority representation from employers. The funds were financed through taxes on profits and government funds, with the obligation to invest in Swedish firms in the risk market. By 1990, the funds owned about 5% of the total stock market value, and were each similar in size to medium-sized private institutional shareholders. In order to achieve *partial* socialization of the means of production, wage-earner funds were supposed to gradually transfer title of ownership to trade unions (not to the state). Despite the success of the trial period, the funds were abolished in 1991 by the Conservative government (Marshall 1995:207).

The unemployment rate remained below 3% until the late 1980s, but according to Mike Marshall (1995), the unity of interest began to fade away in the 1970s and 1980s with the emergence of white collar unions opposing the reduction of wage differentials between high and low profit industries; as well as a move towards local bargaining.

This was followed by the dismantling of the wage-earner funds by the 1991 Conservative government, the failure of the labor movement to build the necessary political support for the project of ‘socialization of investment’, as well as the relocation of many Swedish firms now free to move elsewhere in search of lower labor costs and higher profits (Marshall 1995:209). Unemployment reached a record high 9.6% in July 1993, and has since fluctuated between 4 and 7 percent—a rather unrecognizable range for the Swedish full employment tradition.

### **Contemporary ELR Proposals**

It was Hyman P. Minsky (1965, 1966, and 1986) who revived the ELR idea in the mid-sixties and continued to promote it for the next three decades. Other ELR advocates include: Mathew Forstater (1999, 2002), Wendell Gordon (1997), Philip Harvey (1989), Jan Kregel (1991, 1999), Raymond Majewski (2004), William Mitchell (1997, 1998, 2005), Edward Nell (1988), Dimitri Papadimitriou (1998), William Vickrey (1992) and L. Randall Wray (1998) among others.

From the mid-nineties to the present, seminars, workshops, and research funding continued to materialize. Beginning at the Jerome Levy Economics Institute at Bard College in New York, and continuing at the Center for Full Employment and Price Stability (C-FEPS) in Kansas City (Missouri) and at the Center for Full Employment and Equity (CofFEE) in Newcastle (Australia), the ELR model developed into a more rigorous policy proposal drawing from previous ELR ideas and addressing issues pertinent to contemporary economic conditions as well as political concerns. To begin with, the

“last-resort” part was eventually dropped from the name due to its negative connotation. Names such as Public Service Employment (PSE), Buffer Stock Employment (BSE), or Job Guarantee (JG) are now preferred to ELR. Although most of the work has been geared towards advanced economies like the U.S. and Australia, the gist of the ELR model could be adapted to other countries and adjusted to accommodate their institutional characteristics (e.g. Argentina, India).

According to Minsky, ELR can create “an infinitely elastic demand for labor at a floor or minimum wage that does not depend upon long- and short-run profit expectations of business. Since only government can divorce the offering of employment from the profitability of hiring workers, the infinitely elastic demand for labor must be created by government” (Minsky 1986:308).

In the C-FEPS/CofFEE version of ELR, the government *guarantees* a real job opportunity for anyone ready, willing, and able to work at a fixed socially-established basic wage (plus benefits), thus exogenously setting the price of labor. With ELR, the government will provide a price anchor and establish greater price stability. During a recession, the size of the ELR pool increases to absorb workers displaced from the private sector, and when the economy booms it automatically shrinks when ELR workers find employment in the private sector, hence it operates as a buffer stock employment program. The ELR wage is fixed while the quantity of labor in the buffer stock fluctuates. Private sector employers can obtain labor at a mark-up over the ELR fixed wage; hence the price-stabilization feature of the program. Furthermore, ELR reduces the depreciation of skills caused by unemployment, it contains training

component to prepare participants for private sector employment. ELR also gives more opportunities and freedom of choice for both workers and employers. Participation in the program is voluntary, and it is not a “make work” program. In addition, ELR does not displace private sector jobs since it offers jobs which are undersupplied or not supplied at all by the private sector; those include: companions to the elderly, public school classroom assistants, safety monitors, low-income housing restoration engineers, environmental safety monitors, daycare assistants for ELR workers, community and cultural historians, and ELR artists or musicians (Wray 1998:122-54).

Building on Lerner’s functional finance theory, ELR advocates argue that the government always has the financial capacity to pay for the program. Unemployment only develops “because government spending is insufficient relative to private savings” (Mitchell 2001:23). The size of the deficit necessary to maintain full employment is irrelevant; and so is the national debt for the simple reason that the logic of government finances is totally different from that of households or firms. ELR proponents show that tax payments *do not* and *can not* finance government spending; for at the aggregate level, only the government can be the “net” supplier of fiat money. As a result, the starting point is government expenditure. Once government spends (creates or supplies) fiat money to purchase goods and services, it provides the private sector with the necessary amount of money to meet tax liabilities, save, and maintain transaction balances. The government can safely run a deficit up to the point where it has provided the quantity of non-interest-earning fiat money and interest-

earning bonds desired by the public (Wray 1998).

ELR critics often claim that the program would increase labor bargaining power since it eliminates the threat of unemployment; thus putting more pressure on the wage-inflation spiral. In response, ELR supporters argue that a *skilled* pool of *employable* ELR workers presents a greater “threat” to private sector employees than the traditional reserve army of the unemployed. Thus, one should not expect runaway inflation to develop under the ELR program (Wray 1998). Furthermore, the additional amount of government spending can hardly be inflationary given the low cost of running the program. Critics, however, argue that the inflationary outcome will eventually depend on whether ELR workers produce sealable or non-sealable output.

Estimates for the U.S., U.K. and Australia have shown that the cost of financing ELR ranges between less than 1% of GDP for the U.S. to about 3.5 of GDP in Australia (Mitchell and Watts 1997, Gordon 1997, Kitson *et al.* 1997, Majewski 2004). However, these estimates overstate the real cost of financing the program because they ignore the multiplier effects generated by the new income earned by ELR workers. ELR proponents also argue that the program will pay for itself through the reduction in other social spending associated with unemployment (unemployment benefits, food stamps, crime, police and courts, etc).

Gordon correctly concludes that “beyond this, there is an important sense in which the job guarantee program would not cost anything. The goods or services produced by the labor of the beneficiary of the job guarantee increase the gross national product and the national welfare by as much as the worker is paid as reliably as does any ‘free

market' labor. The laborer is 'earning' the wage or salary received. Also, and importantly, the worker under the job guarantee program has a job of which the worker can be as proud as are other citizens with their jobs" (Gordon 1997:831).

### **Challenges Facing ELR**

Most ELR critics generally support the idea of full employment but have doubts about its capability to deal with structural unemployment, inflation, and logistical problems. Many still remain skeptical about the economic usefulness of the jobs to be created under ELR and the transferability of the skills learned under the program to the private sector (Sawyer 2003,2005; Kadmos and O'Hara 2000; King 2000).

Structural and Technological Change (STC) is a constant feature of capitalist economies. Currently, however, governments do not have any systematic way of dealing with STC. At best, they react to STC after it has happened and after workers have been displaced, only then does retraining begin and the search for solutions is undertaken. Governments only deal with the problem after it has made appearance on the surface instead of dealing with its root causes. This is a very inefficient and irresponsible way of dealing with STC. ELR does have the potential to target STC through careful planning in cooperation with business and union leaders as well as technical training experts in order to constantly study the structural changes in the economy. This allows ELR to stand ready to provide job training for displaced workers so that they can reintegrate the labor force in the most effective way. ELR can therefore provide a systematic preventive program to minimize the damage caused by STC. STC is an institutional problem; therefore the

solution for it must be institutionalized as well. ELR must have a watch list of at-risk industries and at-risk regions so that the ELR administration can stand ready to provide ELR jobs and retraining programs in the areas affected and for the skills needed. Laws can also be introduced to make it mandatory for at-risk industries to alert ELR authorities of imminent closures so that ELR jobs can be planned accordingly.

The inflation threat suggested by ELR critics stems from the claim that ELR is nothing but a Keynesian aggregate demand stimulus policy and is equivalent to pump priming. With ELR in place, however, full employment is guaranteed regardless of the level of aggregate demand. Furthermore, ELR is to be financed like any other government program, by crediting bank accounts, not by "borrowing" or "printing money" (taxes are collected by debiting bank accounts), so ELR spending will simply increase bank reserves (Bell 2000; Mosler 1997-98).

Since ELR workers would receive a living wage, this is often viewed as a trigger for a wage-price inflationary spiral given that all workers receiving less than the ELR wage will demand higher wages (and similar working conditions and benefits). This will cause a one-time (desirable) wage adjustment across the economy that might be accompanied by a one-time fall in profits and does not have to create accelerating inflation. Under ELR, the central bank cannot fight inflation by raising interest rates (to create unemployment) because this would merely increase the size of the ELR pool and the size of the government. The appropriate central bank policy would be open market operations to maintain the short term interest rate at the desired target. In addition, it is noteworthy to mention that all



estimates indicate that the cost of financing ELR (including logistics) is too small relative to GDP to be considered inflationary.

### **ELR at Work: Argentina's *Plan Jefes de Hogar***

After a decade of strict orthodox policies ranging from adopting a currency board and opening markets to foreign trade, to downsizing government and freeing capital; Argentina's economy collapsed, pushing unemployment above 20%. Consumer inflation reached 40% while producer prices skyrocketed by 125%, GDP fell dramatically and the peso depreciated by more than 200%. It was under these dire conditions that the *Plan Jefes de Hogar* (Head-of-Household Program, *Jefes* henceforth) was born in January 2002 via presidential decree during the short term of President Eduardo Duhalde, and came into effect in April 2002. *Jefes* was essentially inspired by the C-FEPS/CofFEE ELR model, but unlike ELR which guarantees employment for all; *Jefes* limits participation to the heads of households which contain children under age 18, persons with handicaps, or a pregnant woman.

The program provides a payment of 150 pesos per month to the head of household for a minimum of four hours of work daily. *Jefes* workers participate in community services and small construction or maintenance activities, or are directed to training programs (including finishing basic education). 87% of *Jefes* beneficiaries work in community projects, including primarily agricultural micro-enterprises and various social and community services such as: cleaning and environmental support in the agricultural sector, improving the sewer systems and water-drainages. Large-scale

infrastructure projects, primarily under the jurisdiction of the Ministry of Infrastructure, also hire *Jefes* workers for the repair of Argentina's roads and bridges.

In 2002, only four months of the implementation of *Jefes*, the indigence rates among participating households had fallen by nearly 25% and among individuals by over 18%. The government finances no more than 80% of the various *Jefes* projects. This provision requires that firms and NGOs executing *Jefes* projects contribute with their own resources. In 2005, Argentina's total government spending on *Jefes* reached about 1.6 million pesos (less than 1% of GDP), which is a price worth paying for a country that has a national poverty rate above 50%, 9.6 million indigents, and a child poverty rate close to 75% (Wray and Tcherneva 2005).

A major concern for ELR advocates is that the *Jefes* program is financed through a World Bank loan in dollars and therefore it is impossible that the community projects designed to improve the living conditions of the poorest can generate dollars for repayment. Recent developments, however, show that the World Bank loans were actually used to repay foreign debt (not to finance *Jefes*). In December 2005, Argentina announced that it will pay off its IMF debt ahead of schedule and stop borrowing from international institutions, while at the same time continuing to implement *Jefes*.

*Jefes* has also increased income of poor households, although it has not pulled them above the poverty line because the program restricts participation to heads of household and because the income it provides is below the official poverty line. Thus, ELR proponents argue that *Jefes* is just a step in the right direction and that it has to 1) be extended to allow participation of anyone

ready, willing, and able to work; 2) pay living wages; and 3) increase its education and training component to meet current needs (Tcherneva and Wray 2005).

There has been a large influx of women into the program who previously were outside the labor force. Women account for over 60% of program participants. The Argentine Ministry of Labor reports that there has been mobility from *Jefes* into the private sector, and that the program has an overwhelmingly positive impact on growth with an estimated multiplier effect of 2.57.

Tcherneva and Wray (2005) conclude that “the program has been a tremendous success, providing jobs to 2 million workers or about 5% of the population, and about 13% of the labor force,” and that despite the huge size of the program, local communities did not experience any shortage of ideas to find useful work for *Jefes* participants. The program is decentralized (operated in coordination with municipalities, NGOs, and non-profit organizations), which has increased political participation and fostered grass-root democracy among traditionally marginalized groups.

### **ELR in India: National Rural Employment Guarantee**

On 5 September 2005, the Indian Parliament passed the National Rural Employment Guarantee Act (NREGA 2005). The law guarantees 100 days per year of employment on rural public works projects to a member of every household in 200 of India’s 600 districts, a scheme which is to be extended to all other districts within five years. The economic architect of NREGA 2005 is Jean Drèze from the National Advisory Council. Drèze (2004) estimates that guaranteeing 100 days of employment per poor household in India will cost Rs 40,000 crores per year

at 2004-05 prices, or 1.3% of GDP. Once the program is phased in, it is expected that the number of poor households will decrease and with GDP rising, the program could be revised such as to guarantee employment to every adult instead of every household, or to increase the cap to more than 100 days per year. The program is targeted at labor-intensive work in the field of environmental conservation and restoration, involving asset-creating public works such as watershed development, land regeneration, prevention of soil erosion, and restoration of tanks.

NREGA 2005 is a first step towards a full-fledged ELR program. It has been designed for the specific needs of India and it will be carefully phased in and merged with the preexisting public works programs over the next five years. It is expected that rural poverty will diminish significantly and that land productivity and environmental conditions will be enhanced as well. The Indian experience shows that ELR schemes are not exclusively for rich countries, and that developing countries are also capable of implementing full employment policies.

### **ELR in France: Professional Transition Contracts**

It was in 1984 (right before the 1986 elections) that Jacques Attali, first put forward his ELR plan to President François Mitterrand who liked the idea but was occupied with other matters and never followed up on it. Attali’s ELR plan reemerged again in newspaper articles in 1994 and 2004. It was the latter that finally got people’s attention and is now being seriously considered for adoption. The gist of the proposed program stems from the idea that unemployed persons who are actively seeking work or actively engaged in training

and skill improving activities outside the labor market are performing a socially useful activity. They deserve to be remunerated for their activity instead of being excluded from social benefits and punished for not finding a job despite their serious job-search efforts.

In December 2005, Prime Minister Dominique de Villepin announced that a pilot experiment of the proposed ELR program will be conducted in six districts. The experimental program will be evaluated in the first quarter of 2007 before being officially adopted nationwide. Workers laid off from companies of less than 300 employees will be eligible for a “Professional Transition Contract” (*contrat de transition professionnelle: CTP*). The CTP workers will sign a contract with a government agency which will guarantee practically the same remuneration as their prior employment. CTP workers will be employed in private companies or public organizations.

The program not only guarantees “activity income” for those who are actively seeking work, but will also provide “individualized coaching” and follow up for the passive unemployment category, in addition to a job training component to facilitate mobility to new occupations due to structural and technological changes.

The CTP contract will be financed by unemployment insurance and that companies who use the services of CPT workers; and if need be the government will cover the remaining cost. The total cost of the program is estimated at 70 billion euros or 4% of GDP (3% activity income, .4% training programs, and .6% individualized follow up) if all the unemployed were to enter the program immediately (Attali and Champain 2005). This would be a less expensive program than the current 4.2% of GDP spent

on unemployment compensations and other employment programs. Like other ELR schemes, the CTP program is a full employment policy that does not increase the deficit and national debt nor does it create inflationary pressures since the unemployment insurance is totally financed by workers and employers.

### **Giving ELR a Chance**

Especially after the Great Depression, there has never been a shortage ELR schemes. Historically, when given a chance, ELR schemes were successful but business opposition, labor disorganization, lack of political support, and misunderstanding of the working of government finances proved to be the greatest obstacles to maintaining full employment in capitalist societies. Full Employment policies in the U.S. (New Deal period), Sweden (Corporatism), and most recently in Argentina (*Jefes de Hogar*), show that ELR schemes can deliver high employment levels without inducing accelerating inflation. India and France are today on their way to implementing ELR schemes designed to fit their specific institutional characteristics. Like any other policy, ELR might have logistical problems but its social, political, economic, and environmental benefits by far outweigh its costs (financial or otherwise). All countries have some sort of public service employment schemes that operate as ELR schemes, but those are so limited in size and scope that their impact is minimal due to self-imposed constraints and/or Washington Consensus policy constraints.

### **Internet Sites**

Centre for Full Employment and Price Stability. [www.cfeps.org](http://www.cfeps.org)

Centre for Full Employment and Equity  
<http://e1.newcastle.edu.au/coffee>  
 Activity, Employment and Job Search.  
[www.supprimerlechomage.org](http://www.supprimerlechomage.org)  
 Right to Food Campaign.  
[www.righttofoodindia.org](http://www.righttofoodindia.org)

## Selected References

- Attali, Jacques; and Champaign, Vincent. (2005) "Activity, Employment and the Search for Employment: Changing the Paradigm to Eliminate Unemployment," *Fondation Jean Jaures*.
- Bell, Stephanie. (2000) "Do Taxes and Bonds Finance Government Spending?" *Journal of Economic Issues*, Volume 34, Number 3, pp. 603-620.
- Beveridge, William H. (1945) *Full Employment in a Free Society*. New York: W.W. Norton & Company, Inc.
- Drèze, Jean. (2004) *Financial Implications of an Employment Guarantee Act Preliminary Estimates*. Note prepared on behalf of the National Advisory Council (August 15).
- Forstater, Mathew. (2002) "Full Employment Policies Must Consider Effective Demand and Structural and Technological Change: A Prime Point of Pasinetti's Political Economy", in Paul Davidson (Editor) *A Post Keynesian Perspective on 21<sup>st</sup> Century Economic Problems*. Cheltenham, UK & Northampton, MA: Edward Elgar.
- Forstater, Mathew. (1999) "Functional Finance and Full Employment: Lessons from Lerner for Today," *Journal of Economic Issues*, Volume 33, Number 2, pp. 475-482.
- Forstater, Mathew. (1998) "Flexible Full Employment: Structural Implications of Discretionary Public Sector Employment," *Journal of Economic Issues*, Volume 32, Number 2, pp. 557-564.
- Ginsburg, Helen. (1983) *Full Employment and Public Policy: The United States and Sweden*. Lexington, MA: Lexington Books.
- Gordon, Wendell. (1997) "Job Assurance: The Job Guarantee Revisited," *Journal of Economic Issues*, Volume 32, Number 3, pp. 826-834.
- Harvey, Philip. (1989) *Securing the Right to Employment*. Princeton, NJ: Princeton University Press.
- Kadmos, George Anthony and Phillip Anthony O'Hara. (2000) "The Taxes-Drive-Money and Employer of Last Resort Approach to Government Policy," *Journal of Economic and Social Policy*, Volume 5 (Summer), pp. 1-22.
- King, John. (2000) "Comment on: George Kadmos and Phillip O'Hara—"The Taxes-Drive-Money and Employer of Last Resort Approach to Government Policy", *Journal of Economic and Social Policy*, Volume 5 (Summer), pp. 72-76.
- Kitson, M., J. Michie and H. Sutherland. (1997) "A Price Well Worth Paying? The Benefits of a Full-Employment Strategy," in Jonathan Michie and John G. Smith (Editors), *Employment and Economic Performance: Jobs, Inflation and Growth*. Oxford: Oxford University Press.
- Kregel, Jan A. (1999) "Price Stability and Full Employment as Complements in a New Europe: A Market-based Price Stabilization Policy for the New ECB," in Paul Davidson & Jan Kregel (Editors), *Full Employment and Price Stability in a Global Economy*. Northampton, MA: Edward Elgar Publishing.
- Kregel, Jan A. (1991) "Is Keynesian Full Employment Policy for 'Top Countries' Only?" in C. de Neubourg (Editor) *The*

- Art of Full Employment: Unemployment Policy in Open Economies*. Amsterdam: North-Holland.
- Lerner, Abba P. (1947) "Money as a Creature of the State", *American Economic Review*, Volume 37 (May), pp. 312-17.
- Lerner, Abba P. (1943) "Functional Finance and the Federal Debt," *Social Research* Volume 10, (February), pp. 38-51.
- Majewski, Raymond. (2004) "Simulating an Employer of Last Resort Program," in George Argyrous, Mat Forstater, and Gary Mongiovi (Editors), *Growth, Distribution, and Effective Demand: Alternatives to Economic Orthodoxy, Essays in Honor of Edward J. Nell*. Armonk, NY: M.E. Sharpe Inc.
- Marshall, Mike. (1995) "Lessons from the Experience of the Swedish Model," in Philip Arestis & Michael Marshall (Editors), *The Political Economy of Full Employment: Conservatism, Corporatism and Institutional Change*. Cheltenham, U.K. & Northampton, MA: Edward Elgar Publishing.
- Minsky, Hyman P. (1986) *Stabilizing an Unstable Economy*. Princeton: Princeton University Press.
- Minsky, Hyman P. (1966) "Tight Full Employment: Let's Heat up the Economy," in H.P. Miller (Editor), *Poverty American Style*. Belmont, CA: Wadsworth.
- Minsky, Hyman P. (1965) "The Role of Employment Policy," in M.S. Gordon (Editor), *Poverty in America*. San Francisco, CA: Chandler.
- Mitchell, William F.; and Wray, L. Randall (2005) "In Defense of Employer of Last Resort: A Response to Malcolm Sawyer," *Journal of Economic Issues*, Volume 39, Number 1, pp. 235-244.
- Mitchell, William F. (1998) "The Buffer Stock Employment Model and the NAIRU: The Path to Full Employment," *Journal of Economic Issues*, Volume 32, Number 2, pp. 547-556.
- Mitchell, William F. and Martin J. Watts. (1997) "The Path to Full Employment", *Australian Economic Review*, Volume 30 (December), pp. 433-435.
- Mosler, Warren B. (1997-98) "Full Employment and Price Stability", *Journal of Post Keynesian Economics*, Volume 20 (Winter), pp. 167-182.
- Nell, Edward J. (1988) *Prosperity and Public Spending: Transformational Growth and the Role of Government*. Boston, MA: Unwin Hyman.
- Papadimitriou, Dimitri B., Mathew Forstater and Wray, L. Randall. (1998) "Toward Full Employment Without Inflation: The Job Opportunity Program", *Report*, Jerome Levy Economics Institute of Bard College, Volume 8, Number 3, pp. 5-8.
- Pierson, John H. G. (1980) *Full Employment without Inflation: Essays on the Economic Performance Insurance (EPI) Proposal*. Montclair, NJ: Allanheld, Osmun & Co. Publishers, Inc.
- Pierson, John H.G. (1964) *Insuring Full Employment: A United States Policy for Domestic Prosperity and World Development*. New York: Viking Press.
- Pierson, John H.G. (1941) *Full Employment*, New Haven, CT: Yale University Press.
- Petty, William. (1662) *The Economic Writings of Sir William Petty, Vol. 1*. Cambridge University Press, 1999.
- Rochon, Louis-Philippe; and Matias Vernengo. (2003) "State Money and the Real World: Or Chartalism and its Discontents", *Journal of Post Keynesian Economics*, Volume 26 (Fall), pp. 57-67.

Sawyer, Malcolm. (2005) "Employer of Last Resort: A Reply to My Critics," *Journal of Economic Issues*, Volume 39, Number 1, pp. 256-264.

Sawyer, Malcolm. (2003) "Employer of Last Resort: Could It Deliver Full Employment and Price Stability?" *Journal of Economic Issues*, Volume 37, Number 4, pp. 881-907.

Tcherneva, Pavlina and. L. Randall Wray. (2005) "Is Argentina's *Jefes de Hogar* an Employer of Last Resort Program?" *Center for Full Employment and Price Stability*. University of Missouri, Kansas City. Working Paper #43 (August).  
[www.cfeps.org/pubs/](http://www.cfeps.org/pubs/)

Tcherneva, Pavlina; and. Wray, L. Randall. (2005) "Employer of Last Resort: A Case Study of Argentina's *Jefes* Program" *Center for Full Employment and Price Stability*. University of Missouri, Kansas City. Working Paper #41 (April).  
[www.cfeps.org/pubs/](http://www.cfeps.org/pubs/)

Vickrey, William S. (1992) "Chock-Full Employment without Increased Inflation," *American Economic Review*, Volume 82 (May), pp. 341-345.

Wernette, John Philip. (1945) *Financing Full Employment*. Cambridge, MA: Harvard University Press.

Wray, L. Randall. (2004) (Editor) *Credit and State Theories of Money: The Contributions of A. Mitchell Innes*. Northampton, MA: Edward Elgar Publishing.

Wray, L. Randall. (1998) *Understanding Modern Money: The Key to Full Employment and Price Stability*. Northampton, MA: Edward Elgar Publishing.

*Fadhel Kaboub*  
*Department of Economics*  
*Denison University*  
*Granville, Ohio, USA*  
*kaboubf@denison.edu*



## Employment Agencies and Temporary Work

*Julia Connell*

### **Introduction**

Over the past 20 years the rise in atypical employment, particularly temporary work, has increased dramatically in developed countries. In the US, the temporary employment agency business is the second fastest-growing industry, with approximately 2.8 million American workers operating as temps on any given day (Alternatives 2001). In China, it has been estimated that following the privatisation of state enterprises the number of temporary workers rose to 70 million in 2003 (Ding 2004; Cheng 2004; Lee 2004). In Korea, the share of temporary workers increased from 43 per cent prior to 1997 to approximately 52 per cent in the post-1997 period (Park, 2004). In Japan, there were over 14 million irregular employees in 2002 (Goka & Sato 2004) whereas Singapore was rather slower than the US and the EU to pick up on contingent employment. In Singapore, the number of temporary or contract workers rose to 37,000 in 2003—a 50% jump from 1998 (Lee 2003). Predictions are that by 2010 one in three Australian workers will be temporary workers and numbers are already close to that in 2004 (ACTU 2003). Consequently, it is not surprising that a public policy report on young adults predicted that if the current growth rates for the temp industry continue, one in six young workers will be employed by a temp agency before turning 35 years of age (Jorgensen, 1999).

Temporary agency working (temping) involves a triangular arrangement in which a temporary work agency (TWA) hires a

worker for the purpose of placing him or her at the disposal of a third party, the user enterprise, for a temporary assignment (Bronstein, 1991:292). Temping is now a fundamental component of the labour use practices for many public and private business enterprises. One of the major drivers of this increase relates to the demand for labour flexibility from user firms and the capabilities of supply from temporary work agencies (see Hall 2002; EU 2002; Storrie 2002).

Increasing demand for labour flexibility has, in turn, resulted from the externalisation of non-core operations by many firms over the past 30 years (ILO 1994:9). Subsequent fluctuating and uncertain production and service requirements have led to a dramatic rise in agency working at an international level. This increase is reflected by the labour hire firms Adecco and Manpower, claiming to be the largest private employers in the world, with approximately 2.1 million temporary workers worldwide employed by Manpower during 1999 (Storrie 2002:27). Manpower originated in the US in 1948, opened its first offices in the UK in 1956, and in France in 1957. Hence, the use of temporary work agencies specialising in the supply and provision of workers to client companies dates back to at least the 1950s.

Having briefly explored the topic of TWAs and the extent of temporary working in a range of countries, this chapter now examines the definition of temporary work and TWAs, what TWAs actually do, the growth of the TWA sector, before examining a number of policy implications arising from these issues.

### **Temporary Work - Key Definitions**

The OECD (1996) notes that definitions of temporary work differ across countries

according to relevant systems of employment regulation. There is, however, generally a key distinction between TWAs and what are commonly referred to as employment brokers and employment placement agencies (Gonos 1997). Although employment brokers are involved in triangular relationships (in common with TWAs) the substance of the relationship differs. As Campbell, Watson and Buchanan (2004) point out, an employment broker alerts job seekers to employment vacancies, screens potential recruits for the employer, accepts a fee for its service, and then *withdraws from the scene*. By contrast, a TWA enters two contracts: one with the temp worker for the use of their labour and another with a user (or host) company to supply labour services. The temporary work agency thereby remains in an *ongoing relationship* with both parties.

In Taiwan and Japan agency temporary workers are known as ‘dispatched workers’. In Japan they participate in ‘indirect relationships’ between the dispatching agency, the client and the dispatched worker. The employment agency and the client company make a contract with each other with the latter paying the former a fee for ‘dispatching’ the worker. Nonetheless, the employment relationship exists *only* between the dispatching agency and the dispatched worker, even though the latter works for the client company. Thus, the client company does not enter the employment relationship, but directs and supervises the dispatched worker without actually hiring him or her (Goka and Sato 2004).

Despite regulatory differences between countries, there appear to be a number of distinctive working arrangements that can be regarded as being typical of temporary agency employment. Specifically, the main

features of temporary agency employment are that contracts are for limited duration (that is, they are insecure); the employment relationship involves three parties; and simultaneous employment engagements (multiple job holding) may be associated with such employment.

Frequently, the distinction between temporary agency work and employment broking is not clear, and consequently, has formed the basis of extensive debate in many countries. This is partly because temporary agency work in its modern form arose out of the activities of employment brokers (Campbell, Watson and Buchanan 2004). Nevertheless, it is important to make the distinction, because it underlies the major features of temporary agency work and has significant implications for its growth. Whereas, employment brokers operate in specific occupational labour markets, where they play an intermediary role within labour market transactions, temporary work agencies intercede in the employment relationship itself, dividing up and redistributing the managerial practices that were previously part of the employment relationship. Some practices remain with the user firm, which retains the responsibility for directing labour and some are transferred to the TWA, which commonly now claims the role of legal employer.

These definitions and distinctions apply to temporary agency work in most advanced societies. Yet if the relationship between the TWA and the temp worker is examined some significant cross-national differences can be determined. For example, given the basic definitions of temporary agency work (Michon 1999; Storrie 2002), it is assumed that the worker is an employee of the temporary work agency. This corresponds with regulation in most countries (Storrie



2002). In Australia, however, regulation is unclear, and as a result, so is the employment status of temporary agency workers. Some workers organized by temporary work agencies are accepted as employees of the agency, but others are often regarded as contractors; that is, as self-employed workers.

### **Temporary Work Contracts and Practices**

According to Storrie (2002), temporary agency working is not included in the employment typology in many EU countries and many national authorities responsible for gathering labour market statistics have yet to include it in their national surveys. For example, retirees may be on the books of temp agencies and prepared to carry out the occasional short-term engagement, since they possess specialist skills. The unemployed may sign on with temp agencies in order to obtain work experience, training and job placement such as occurs under the umbrella of the Job Network in Australia. Indeed, in Australia, it is the provision of labour services to the unemployed that has provided an indirect boost to TWAs.

In Latin America the development of what Riquelme (1987) refers to as 'capitalist agriculture' led to the intensive use of modern technology, new labour systems and new work practices such as the recruitment of women as temporary workers. This 'preferred practice' occurs without contracts and pay is organised through 'piece-work', usually meaning that women have to work longer hours to meet assigned targets than they would have done on regular contracts (Riquelme 1987). De Leal (1987) maintains that the absence of a written contract makes the female worker 'invisible' and, therefore, not included in statistics as verbal contracts

do not appear in company staff records. This also happens in China where employers only need to provide oral agreement concerning the labour contract if employment lasts less than one month (Ding 2004).

Moreover, short-term employment practices associated with labour hire may also be associated with clandestine activities such as the employment of illegal immigrants and tax avoidance. As such, there will be a degree of 'under reporting' of total employment in the sector. The contracting organisation may itself be an intermediary, so the employment relationship can be further removed through subsequent sub-contracting. Multiple employer and multiple contracting arrangements serve to further complicate the employment relationship, the status of the agency worker and the responsibilities of the parties involved (Rubery *et al.* 2000). In addition, an individual temp agency worker may also shift between assignments across several agencies.

### **What Do Temporary Work Agencies Do?**

The temporary work industry in many areas is largely unregulated with relatively free entry and exit. This makes it difficult to track its development. Nonetheless, it is evident that one of the driving forces for the increase in temporary working internationally has been the demand from user firms and the ability to supply from temporary work agencies (TWAs). One of the most frequently cited reasons for user firm demand is the labour flexibility that temp contracts provide (Carre & Tilly 1998; Connell & Burgess 2002; Gonos 1997). Hence, TWAs become purveyors of flexibility and HR 'brokers', frequently undertaking the HR roles that were previously conducted 'in-house' such as

recruitment and selection, payroll, appraisal and training.

The NSW Labour Hire Taskforce (2001:19) cites the main services provided by labour hire companies as including: supplementary labour (short term hires, usually on an hourly basis; managed services); the provision of outsourced services on a project basis; direct contract arrangements (placement of individuals into contract arrangements; and recruitment services).

Many TWAs also offer specialist services to particular industries (for example, mining and health) or they specialise in particular occupations (such as clerical temps). Apart from the large multinational temp agencies there are also a large number of small agencies that are generally owner-operated.

There are signs that temporary agency services can go beyond limited assignments, such as purely placing temps with user firms. Further developments involve longer-term relationships between the TWAs and user firms with TWAs taking on more and more managerial functions. Peck and Theodore (1998:663-665,668) stress this development in their study of temporary work agencies in Chicago. They argue that many agencies are doing far more than delivering 'warm bodies' by moving into long-term, human-resource-based functions.

The functions undertaken by TWAs can also extend to 'on-site arrangements' for employers, in which the agencies contract to service and manage a complete job function, including processes of supervision and performance monitoring. Peck and Theodore (1998:665) describe this in terms of temp agencies deepening and broadening their sphere of control "from the labour market to the labour process: from the predominantly labour-market practices of hiring, screening,

(re)allocation, scheduling and payment of workers, through to the predominately labour-process practices of induction and training, supervision, task allocation, performance monitoring and relations with other (temp and non-temp) workers".

Campbell et al (2004) point out that these shifts in responsibility pose major consequences for workers, since they threaten to break up the standard form of direct employment within individual firms, together with all the rights and benefits that have come to be associated with this form of employment – ultimately resulting in a redesign of the employment relationship.

### **Growth of the Temporary Work Sector**

Globally, the temporary agency industry features a number of very large, high profile labour hire, or 'flexible labour' firms that have moved in 'over the top' of domestic operators, on occasion acquiring ownership and control of those companies. Currently Adecco and Manpower are the two largest labour hire operators worldwide (Rivito, 2004) with Manpower owning or franchising 4,300 offices in nearly 70 countries (mainly France, the UK, and the US) up from 2,400 in 43 countries in 1997 (Schellhardt 1997). Kelly Services, the second largest staffing company in the US in 2000, has also gained a foothold in Asia with 17 offices located in Singapore, Malaysia, Indonesia, the Philippines, Thailand, India and Hong Kong (ZaoBao 2002). Such TWAs operate on a principal of 'one-stop shopping' for any kind of temp staffing, leading a human relations executive to comment "Adecco are acting more like consultants than just a people factory" (Sansoni 1997).

The temporary work sector is also expanding across the OECD as a number of multinational TWAs develop an

international model of labour brokerage providing labour matching services for job seekers and job providers. These TWAs (such as Manpower, Kelly & Drake) provide services for employers that range from payroll administration to staff appraisal systems. As such, they intermediate between the purchasers and providers of labour and can also be used to blur regulatory responsibility, de-unionise workplaces and reduce wage rates (Peck & Theodore 2002). Consequently, TWAs offer the ultimate form of labour commodification, hiring strictly on a 'needs basis' with no attached obligation or commitment.

Despite the rapid growth of the temporary agency sector, in most EU countries it appears to count for around three per cent of total employment or less. Consequently, it would appear to be not that significant in the overall context of employment. See Table 1, below:

*Table 1. Growth of Temporary Work in EU*

Country	Extent (1999) and Growth
Austria	24,777 (0.7%) quadrupled since 1992
Belgium	62,661 (1.65%) doubled since 1992
Denmark	18,639 (0.9%) 5-fold increase since 1992
Finland	15,000 (0.6%) from 11,000 since 1996
France	623,000 (2.7%) rapid growth
Germany	243,000 (0.7%) doubled since 1992
Ireland	9,000 (0.6%) moderate recent growth
Italy	31,000 (1.5%) very rapid recent growth
Luxembourg	6,065 (2.3%) doubled since 1992
Netherlands	305,000 (4%) doubled since 1992
Portugal	45,000 (1%) doubled since 1995
Spain	109,000 (0.8%) 5-fold increase since 1995
Sweden	32,000 (0.8%) rapid current growth
UK	557,000(2.1%) 3-fold increase since 1992

*Source:* Adapted from Storrie (2002)

However, in a review of the EU temporary work sector (Storrie 2002) a number of features emerged in relation to

temporary agency work in Europe. First, there is a chronic data shortage. Second, the regulation of temporary employment agencies is minimal, with a requirement for annual registration but no provision for financial guarantees and reporting and, third, there is no consistent legislative code in dealing with temporary work, although there are emerging developments.

### **What is Driving Expansion in Agency Employment?**

As Burgess and Connell (2004) point out, the rationale for hiring temporary workers tends to be the same as many other workplace initiatives – labour cost savings associated with downsizing, increased global competition, the introduction of new technology and the need to respond quickly to an ever-changing marketplace. In Japan, Goka and Sato (2004) comment that the employers of large companies expect the employment of dispatched workers (temps) to eliminate the need for high salaried, middle-aged employees whereas the government and business community expect temp working to decrease the number of unemployed.

The rise of the temporary agency sector can also be interpreted as one manifestation of increased flexibility in which the restructuring of internal labour markets and lower internal labour costs ('headcount costs') are associated with organizations externally shifting recruitment, training and 'on-costs' to the temporary agencies and temporary workers. Temporary agency employment not only offers flexibility and cost saving potential, it also potentially removes responsibility for compliance with many employment regulations such as unfair dismissal, employment insurance, employment benefits and superannuation

entitlements onto the TWA. In this context, temping allows for a shifting of the responsibilities and risks associated with direct employer responsibilities. As such there have been claims that organisations are shedding permanent jobs and hiring in agency workers (see Telstra on Charges, Workers Online 2003).

Traditional reasons that are driving the growth of the temp sector include labour shortages in relation to the professions including IT, accounting, nursing and teaching. This is evident in the US, where Peck and Theodore (2004) comment that despite strong growth in the 'core' clerical market for temps, by 1990 it represented less than one third of total temp placements with rapid expansion in 'new' sectors such as healthcare, government markets, and in some technical and professional fields, leading to an increasingly complex and diversified sector.

That said, data from the Recruitment and Consulting Services Association (RCSA) in Australia suggests an over-representation, compared to national employment data, in clerical and blue-collar occupations, and in the mining, manufacturing and construction sectors. This is not surprising since in some industries labour demand is linked to contracts that have a limited duration and/or are linked to a specific task.

In India, temp workers are not covered by labour laws and are said to be mostly low or semi-skilled, carrying out duties such as cleaning and other manual tasks (Das and Dhananjay 2004). In Taiwan, however, as in several other Asian countries temps can be quite highly skilled working as programmers, project managers and similar (Cheng 2004).

Overall, it seems apparent that there is no shortage of short-term temporary

assignments available for those with the requisite skills. For those with family caring responsibilities, semi-retirees, post-graduate students and those who do not require the commitment of a full-time and ongoing job, temping can satisfy life-style options and complement non-work activities. That is, the demand imperative for more just-in-time employment arrangements is being matched by the global growth of an industry that can facilitate this process. As already discussed, this particular employment arrangement is characterised by ambiguity in terms of employment status and in terms of employer responsibilities. Burgess and Connell (2004) argue that for some occupations (for example nursing, teachers, IT programmers) the shortage of trained and available workers is driving the expansion of TWAs, but, as argued previously, in many cases it is the possibility of lower labour costs and further labour flexibility that is providing employers with the incentive to replace permanent workers with agency workers. This, in turn, will place pressure on employment security and on the employment conditions of permanent workers.

## **Policy Issues**

The situation in Australia is broadly similar to that in the UK and the largely unregulated temporary hire sector in the USA (Peck and Theodore 2001). Specifically, there has been an absence of regulation, there is confusion over status within the temporary work sector and there are no exclusions with respect to temporary agency employment arrangements. Unlike the EU there is no social directive or national policy that attempts to clearly set out the rights associated with labour hire employment.

The growth in temporary agency work has been supported by large transnational

employment agencies that have expanded their business in both the private and public sectors. These developments have a number of important social policy implications. For example, temporary agency workers are liable to have ambiguous employment status, not belong to trade unions and have precarious employment arrangements. This is partly because in many countries the temporary work industry has evolved and developed in a largely unregulated context. Consequently, there are emerging problems in terms of employer responsibility, protection of employment benefits, enforcement of minimum employment conditions, responsibility for occupational health and safety conditions, trade union access and responsibility for training.

In Australia, there is considerable public concern over large corporations restructuring their workforce through the replacement of permanent workers with labour hire workers (Workers Online 2003), for the potential for labour hire workers to be subject to greater risk from workplace accidents (Underhill 2003), for labour hire workers to be without many rights and protections associated with employee status (Hall 2002) and, for labour hire in itself to undermine formal and informal forms of on-the-job training and skills acquisition (Connell & Burgess 2001).

From the literature presented in this chapter, it is evident that temp labour hire represents a gap that can potentially be exploited and lead to nefarious practices that ultimately are likely to undermine the credibility of the temp industry and impose a cost on the public sector. The situation of ambiguous employment status has potentially adverse implications for labour rights, taxation status, access to employment benefits and workers compensation. In turn, there is scope to use labour hire practices to

de-unionise the workplace and undermine employment conditions in order to avoid statutory obligations. At one end of the spectrum, labour hire organisations may be linked to clandestine activities including illegal immigration and tax evasion. Conversely, the reputable end of the industry are said to be seeking more extensive registration and regulatory requirements for temporary work operators (Greenwald 1996).

The continuous tension between the flexibility required by user firms and the protection of temp agency workers makes it necessary to incorporate a public policy dimension within the debate. To date, there has been a surprising lack of regulation surrounding agencies and agency work in some countries while others have introduced policies and regulations which have, in the main, proved to be fairly ineffective. This implies that there are several unanswered public policy issues associated with agency work. For example, what is the 'proper' level of public regulation of agencies and agency work and how can efficient, low cost regulations be implemented and policed? In light of the discussion of the 'knowledge society', it is also important to discuss how vocational education and training can be enhanced and whether there are certain mechanisms and training systems that are better than others? The commercial and academic interest in temp agency work has not been matched to date by a similar interest in the public policy dimension.

## **Conclusion**

Positive outcomes arising from temporary employment are that: it can improve job matching within the labour market, reduce job search time and expenses, and offer a transition to permanent employment (OECD

2001). Temporary work is important for user firms that cannot recruit skilled labour, that are facing product market instability, are engaged in an organisational restructure or have to meet imposed external directives regarding payroll or employee deployment. Nonetheless, there are potential dangers associated with expanding the relative size of the temporary workforce, such as exclusion from standard employment benefits and from an organisation's internal labour market (including training and career paths), job and wage insecurity and segregation into a cycle of contingency employment arrangements (Burgess and Connell 2004). So while some workers, such as the highly paid, executive and professional workers may benefit greatly from temporary working, a second class working group is also being created (Appelbaum et al 2004).

Nevertheless, it is important to see agency work in both a wider theoretical and contextual perspective and provide more specificity about temp agency work. The wider theoretical and contextual perspective is necessary because there is considerable fluidity across countries, local labour markets, industries, occupations and individual people. The impact of context creates a variety of trends which raise a number of questions about people's choices and the consequences of these choices. It also raises issues about so-called work-life issues and what the role of non-work factors is in relation to worker options concerning work and working time. Thus, the complexity of the interplay between situational factors and individual choices needs to be further explored.

The current ambiguity and absence of regulations surrounding the temporary work sector indicates that there is an urgent need

for policy reform with reference to regulatory norms and employment arrangements. As discussed previously in this chapter, although some reforms have been implemented they are frequently either not implemented, difficult to monitor, or ineffective (as per the examples given for China and Latin America. Given that the numbers of temporary workers appears to be steadily increasing it is suggested that policy makers take note and monitor the effectiveness of the recent European Union directive (2002) *European Union on Working Conditions for Temporary Workers* where it is stipulated that temporary workers will no longer be subject to discrimination due to the nature of their employment contract.

### **Selected References**

- Alternatives. (2001) Contingent Work: A Silent Revolution in the American Workforce, *Newsletter of the Center for Policy Alternatives*, Jan-Feb, Volume, Number 1.
- Appelbaum, E., P. Berg and Y. Matsuzuaka. (2004) *Differential Access to Flexible Schedules: Implications for Employee Outcomes*. Paper presented at 2004 IIRA Conference, Seoul, Korea.
- ACTU. (2003) *Future of Work: One In Three Workers Will Be Casual By 2010*. [www.actu.asn.au/public/news/105537624\\_7\\_21338.html](http://www.actu.asn.au/public/news/105537624_7_21338.html)
- Bronstein, A. (1991) "Temporary work in Western Europe: Threat or Complement to Permanent Employment", *International Labour Review*, Volume 130, Number 3, pp. 291-310
- Burgess, J. and J. Connell. (2004) "International Aspects of Temporary Agency Employment: An Overview", in J. Burgess and J. Connell (Editors),

- International Perspectives on Temporary Agency Work*. Routledge: London.
- Campbell, I., I. Watson and J. Buchanan. (2004) "Temporary Agency Work in Australia: Part I", in J. Burgess. and J. Connell (Editors), *International Perspectives on Temporary Agency Work*. Routledge: London.
- Carre, F. and C. Tilly. (1998) "Part-time and Temporary Work: Flexibility for Whom?", *Dollars and Sense*, Jan-Feb No 215, pp. 22-26.
- Cheng, C.Y. (2004) *A Study on the Current Development of Dispatched Work and its impact on Employment Relations in Taiwan*. Paper presented at 2004 IIRA Conference, Seoul, Korea.
- Connell, J and J. Burgess. (2001) "Skills, Training and Workforce Restructuring In Australia", *International Journal of Employment Studies*, Volume 9, Number 1, pp. 1-24.
- Connell, J. and J. Burgess. (2002) "In Search of Flexibility: Implications for Temporary Agency Workers and Human Resource Management", *Australian Bulletin of Labour*, Volume 28, Number 4, pp. 272-283.
- Das, A. and P. Dhananjay. (2004) *Contract Workers in India: Emerging Economic and Social Issues*. Paper presented at 2004 IIRA Conference, Seoul, Korea.
- Ding, S. (2004) Normalized Labor Relations of Temporary Employment, *paper presented at 2004 IIRA Conference, Seoul, Korea*.
- DIR. (Department of Industrial Relations.) (2001) *Labour Hire Task Force Final Report*. Report of the Task Force to the Department of Industrial Relations, New South Wales.
- Easton, D. (1997) *The Commercialisation of New Zealand*. Auckland University Press, Auckland.
- European Union. (2002) *Working Conditions for Temporary Worker*. EU Council Directive, 99/70/EC.
- Goka, K. & Sato, A. (2004) Agency Temporary Work and Government Policy in Contemporary Japan, in J. Burgess & J. Connell (Editors), *International Perspectives on Temporary Work*. Routledge: London, pp 112-128.
- Gonos, G. (1997) "The Contest over "Employer" Status in the Postwar United States: The Case of Temporary Help Firms", *Law and Society Review*, Volume 31, Number 1, pp. 81-110.
- Hall, R. (2002) *Labour Hire in Australia: Motivations Dynamics and Prospects*. Sydney: ACIRRT, University of Sydney
- ILO. (1994) *The Role of Private Employment Agencies in the Functioning of Labour Markets*. Report from the International Labour Conference 81<sup>st</sup> Session. Geneva: International Labour Office.
- Jorgensen, H. (1999) "When Good Jobs Go Bad: Young Adults and Temporary Work in the New Economy." 2030 Center, [www.2030.org/resources/read/asp?ID=87](http://www.2030.org/resources/read/asp?ID=87)
- Lee, J. (2003) Temp Work Thriving in Singapore. [http://new.kellyservices.com.au/www/main.cfm?item\\_id](http://new.kellyservices.com.au/www/main.cfm?item_id)
- Lee, J.S. (2004) Labor and Social Issues in a Global and Dynamic World Economy, Report for Track 3. *Presented at 2004 IIRA Conference, Seoul, Korea*.
- Michon, F. (1999) "Temporary agency work in Europe", *Eironline* January, [www.eiro.eurofound.ie/1999/01study/TN9901201S.html](http://www.eiro.eurofound.ie/1999/01study/TN9901201S.html)
- NSW Labour Hire Task Force. (2001) *Final Report*. Sydney: NSW Department of Industrial Relations.



- OECD. (1996) *Employment Outlook*. Paris: OECD.
- OECD. (2001) *Employment Outlook*. Paris: OECD.
- Park, W., Y. Nho and D. Kim. (2004) Unions and the Use of Contingent Workers: Evidence from Korea. *Paper presented at 2004 IIRA Conference, Seoul, Korea*.
- Peck, J. and N. Theodore. (1998) "The Business of Contingent Work: Growth and Restructuring in Chicago's Temporary Employment Industry", *Work, Employment & Society*, Volume 12, Number 4, pp. :655-674.
- Peck, J. and N. Theodore. (2001) "Contingent Chicago: Restructuring the Spaces of Temporary Labor", *International Journal of Urban and Regional Research*, Volume 25, pp. 471-496.
- Peck, J. and N. Theodore. (2002) "Temped Out? Industry Rhetoric, Labor Regulation, and Economic Restructuring in the Temporary Staffing Business", *Economic and Industrial Democracy*, Volume 23, pp. 143-175.
- Peck, J. and N. Theodore. (2004) "Cycles of Contingency: The Temporary Staffing Industry and Labor Market Adjustment in the U.S.", in J. Burgess and J. Connell (Editors), *International Perspectives on Temporary Agency Work*. Routledge: London. pp 24-36.
- Riquelme, V. (1987) *Review of The Legal Status of Rural Women in nineteen Latin American Countries*.  
[www.fao.org/DOCREP/U5615e/u5615e03.htm](http://www.fao.org/DOCREP/U5615e/u5615e03.htm)
- Rubery, J. Earnshaw, J., Marchington, M., Cooke, F. and Vincent, S. (2000) *Changing Organisational Forms and the Employment Relationship*. Proceedings from the Conference for the Advancement of Socio Economics. London: London School of Economics.
- Sansoni, S. (1997) "Move Over, Manpower", *Forbes*, Volume 160, Number 1, pp. 74-75.
- Schellhardt, T. (1997) "Temporary-Help Industry Now Features Battle of the Giants—Adecco Moves Increase Pressure on Manpower Inc and Smaller Rivals", *Wall Street Journal*, Nov 6, pp. 1
- Spicer, B., D. Emanuel and M. Powell. (1996) *Transforming Government Enterprises*. Melbourne: Centre for Independent Studies.
- Stanworth, C. (2003) Flexibility at a Price—The Case of Office Temps in the UK. *ESC Monograph*. Newcastle, Australia: University of Newcastle, pp. 18-26.
- Storrie, D. (2002) *Temporary Agency Work in the European Union*. Dublin: European Foundation for the Improvement of Living and Working Conditions.
- Workplace Express*. (2000) Adecco Committed to Paying Host Company Rates 19 July. [www.workplaceexpress.com.au/](http://www.workplaceexpress.com.au/)
- Underhill, E. (2003) *Changing Work and OHS: the Challenge of Labour Hire Employment*. [www.actu.asn.au/public/ohs](http://www.actu.asn.au/public/ohs)
- Workers Online (2003) Telstra on Charges. [www.workers.labor.net.au](http://www.workers.labor.net.au)
- ZaoBao, LianHe (2002) *Companies Turn to Hiring Temporary Staff During Downturn*. [new.kellyservices.com.au/www/main.cfm?item\\_id](http://new.kellyservices.com.au/www/main.cfm?item_id)

Julia Connell

Faculty of Business

University of Technology Sydney

Australia

[Julia.Connell@uts.edu.au](mailto:Julia.Connell@uts.edu.au)



## Endogenous and Exogenous Money and Credit

*Sergio Rossi*

### **Introduction**

The nature and role of money and credit never cease to fascinate economists. Since the writings of Aristotle, money has been at centre-stage of economic debate, and many controversies on its origins and functions have been animating the history of monetary thought. The principal questions that are still debated today go to the roots of money's essence in modern economic systems (see Smithin 2000; Rochon & Rossi 2003) and ask:

- What is money?
- How is money created?
- What is the value of money?

Another set of questions, which are ancillary in respect of the above issues, are directed to money management by banks as well as by monetary authorities. To list only a few:

- Can central banks control the money supply?
- What is the causal relationship between money and credit?
- How are interest rates determined and what influence do monetary authorities exert on them?

To answer these, and many other, questions, it is first of all necessary to disentangle the nature of money. Here the problems begin. In fact, as Schumpeter (1954:289) noticed, "views on money are as difficult to describe as are shifting clouds". Despite 200 years of monetary economics, it is no exaggeration to claim that "the definition of money can still be regarded as an almost unresolved issue" (Bofinger 2001:3). Most economists revert indeed to a functional definition of money,

as was suggested by Hicks (1967) and then taken up uncritically by the economics profession, without noticing that this approach is prone to circularity: "If it is not clear what 'money' is, it is also not possible to describe the functions of 'money'" (Bofinger 2001:4). On the other hand, the alternative definition of money, which is based on the statistical properties of some monetary aggregates (M0, M1, M2, M3, and so on), lacks sound theoretical foundations and "puts the cart before the horse" (Osborne 1992:603). Without a precise definition of money, in fact, it is not possible to rely on statistics to draw a conceptual line between money and non-money assets, since no fix point exists on theoretical grounds. Financial innovations add a further problem to this approach in order to define money by monetary aggregation, because they shift the borderline between monetary and non-monetary assets incessantly.

To circumvent these conundrums, there is today a widespread agreement among economists that "[t]he money of monetary theory is the monetary base" (p.606), and a widely-held belief that the monetary base differs from all other forms of money in that it is a completely exogenous variable. To be sure, "[m]any economists regard *exogeneity* as a decisive quality of 'money'" (Bofinger 2001:5). The theoretical consequence of this view, to put it in the words of Withers (1909), is that "deposits make loans", and bank credit functionally depends on (and is limited by) pre-existing loanable funds that banks hold in their vaults (see Realfonzo 1998:ch. 6).

### **Exogenous Money and Credit**

According to the textbook analysis, the money supply is and can be controlled by the central bank, which can affect the

money stock,  $M$ , via the base money multiplier,  $m$ , as epitomised by the formula  $M = mB$ , where  $B$  is the monetary base (also dubbed high-powered money) issued by the central bank. Whilst the textbook story admits that also banks create money—in the form of deposits,  $D$ —it sets off from the hypothesis that the monetary base is an exogenous variable like the reserve/deposit ratio,  $R/D$ , and the currency/deposit ratio,  $C/D$ , both entering the determination of  $m$  (Mankiw 2003:Ch. 18). On this assumption, which is almost as old as the quantity theory of money that is in the backstage and that can be traced back to Hume (1752), the central bank controls the money supply, hence bank credit (see Figure 1), through the determination of the monetary base according to the expected value of the base money multiplier.

Further, in this view the banks' money creation process is closely related to the reserve requirements imposed upon them by the monetary authority, so that their contribution to money creation strictly depends on the rules enforced by monetary policy makers. As a result, the direction of causality runs from bank deposits (reserves) to bank loans, and the rate of interest ( $i^*$ ) is determined by market forces as in the *IS–LM* model.

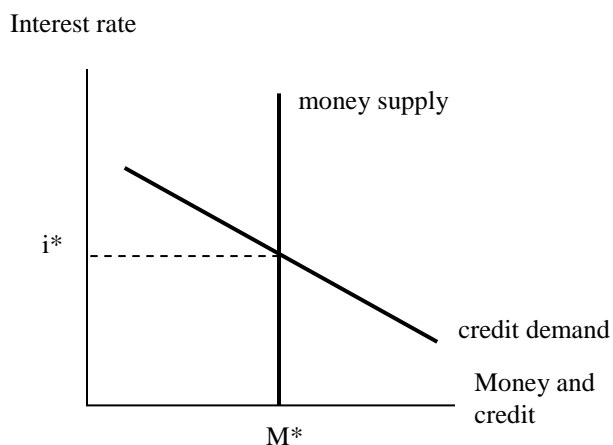


Figure 1. Exogenous Money and Credit

Several problems exist in this framework. First, the nature of (exogenous) money remains to be explained. In fact, the base money multiplier relation is not concerned about what  $M$  is. Neither is the quantity theory of money purporting that  $MV = Py$ , where  $V$  represents the income velocity of money,  $P$  the general price level, and  $y$  real output. To be sure, the quantity theory of money has been developed at a time when the monetary function was exerted using a precious metal like gold or silver. At that time, therefore,  $M$  may have been identified with the stock of commodity-money in circulation, say in the form of gold coins, or with the amount of paper-money representing the gold-money stock under the gold standard system.

Since the demonetisation of gold, and of precious metals in general, however, the money stuff has been dematerialised. This is so much so that today, owing to information technology, it consists of purely electronic impulses in a bank's ledger (Cencini 1995; Ingham 1996). To be sure, modern money is book-entry money, either in the central bank's or in the banks' accounts. If so, then one has to explain why money is accepted in exchange for real goods and services, granted that it is an incorporeal unit created at a trifling cost by the banking system (including here the central bank).

Here lays the second major problem encountered, and unsolved, by adherents to the money exogeneity paradigm. In fact, advocates of exogenous money have always painfully addressed the explanation of money's value, and all their arguments have to be rejected on logical grounds. According to them, the value of money can be explained on practical grounds, since money is the so-called general equivalent

of non-money goods. As Clower (1967:5) puts it in his famous axiom, “money buys goods and goods buy money; but goods do not buy goods”. This, however, is logically not enough to explain why agents use money, an intrinsically valueless ticket *à la* Cannan (1921), “to [enter] the great social store of all goods [that are exchanged against it]” (Schumpeter 1954:289). The explanation put to the fore in various academic quarters that “one person gives up goods (objects that appear as arguments of utility functions, directly or indirectly) for fiat money only because the person believes that someone else will subsequently give up goods for fiat money at an acceptable rate of exchange” (Wallace 1980:49) begs the question. In neoclassical thinking, in fact, any “rate of exchange” is a relative price determined as a result of supply and demand, with or without the friction-avoiding intervention of a Walrasian auctioneer.

This amounts to saying that the value of money can only be established once the relevant exchange has taken place between the real terms of trade,  $a$  and  $b$ , in a market session that involves two poles only (traders A and B). In such a framework, therefore, money is inessential in the sense of Hahn (1973), because it does not serve to determine value and prices, which are based on utility and exchange. The distinction between a monetary and a non-monetary economy would thus be a matter of practice rather than substance: in this framework, the existence of money avoids the practical problem known as the double coincidence of wants, but, essentially, money prices are relative prices that exist in any case as long as the economy consists of at least two distinct real goods, services, or assets that are exchanged in either separate or simultaneous market sessions.

The fact that one of the goods is dubbed “money” does not change the essence of economic transactions, which occur each in a strictly bilateral relationship involving a buyer and a seller of non-money goods.

Further, and closely related to the preceding question, adherents to the exogeneity-of-money view have also to explain how would it be logically possible for money to be, in one and the same transaction, a medium of exchange (that is, a means of payment) and an object of trade supplied and demanded for its own. The problem here is again of a logical nature: it consists in determining if the thing called “money” can logically exert two different functions at one and the same point in time, that is, in one and the same transaction. In particular, logic must establish if money, undoubtedly a means of payment in the real world, is also an object of trade that agents demand and seek to obtain for its own exchange value, and which “plays a distinctive asymmetric role as one side of virtually all transactions” (Starr 1980:263).

In fact, as a general equivalent, money is an object like real goods and services, with its own velocity of circulation, and hence it is the object of supply and demand as are all non-money goods in the economy. In this framework, as Friedman (1987:5) notes, “[m]oney is treated as a stock, not as a flow or a mixture of a flow and a stock”. If so, then money is a liquid store of wealth, to wit, “a temporary abode of purchasing power” in Friedman’s words. This gives rise to the logical problem of giving to money assets a price, even though money is a worthless token used to circulate output. Indeed, as Balasko and Shell (1981:112–13) acknowledge, “[m]oney does not in general serve as a proper store of value—i.e., money cannot have a positive price—in the finite-horizon

economy in which the terminal date is known with certainty. The reason is obvious. Money is worthless at the end of the final period. Consequently, in the next-to-last period, individuals desire to dispose of money holdings in order to avoid capital losses. This drives the price of money to zero at the end of the next-to-last period. And so on. Individuals with foresight drive the price of money to zero in each period, i.e., the ‘general price level’ in equilibrium must be infinite. The natural way to permit money to be a proper store of value is to go beyond the finite-horizon model”. As a result, several money demand formalisations (cash-in-advance models, money-in-the-utility-function models, and money-in-the-production-function models) have been developed over the last three decades, using the portfolio-theoretic approach to asset diversification for finitely-lived “representative” agents within an infinite-horizon framework.

Consider for instance the overlapping-generations (OLG) monetary models that have been developed following Samuelson (1958), and that Wallace (1980:50) regards as the best of the available paradigms for explaining money in modern economic systems. These models analyze the working of our monetary systems assuming an intermediary asset called money, to avert, or at least to reduce, search and bargaining costs among the different generations of traders (usually ranged in two complementary classes, dubbed “young” and “old”). These generations exist in any market period, and they overlap as time goes by. “The OLG model starts by endowing the initial old with the initial stock of money. [...] In period  $t$ , the young individual receives more of the consumption good that he wants to consume but cannot store the excess since

the consumption good is perishable. He sells it to the initial old for fiat money, provided that he expects to be able to exchange his holdings of fiat money for the consumption good in period  $t + 1$ ” (Handa 2000:628–31).

Among the problems raised by this paradigm, such as the origin of the money stock entering the “initial endowments” of the first old generation, as well as the value of money exchanged for the consumption good (see above), the logical problem of distinguishing the medium of exchange from the terms of trade is crucial. As Smith (1776) puts it, in fact, the medium of exchange ought not to be considered as an object itself. Although in Smith’s time money was reified into a precious metal, which blurred the logical distinction between the means of payment and its material support (gold for example), in *An Inquiry into the Nature and Causes of the Wealth of Nations* Smith already pointed out the importance to distinguish money proper from money’s worth, both analytically and in practice. “The great wheel of circulation is altogether different from the goods which are circulated by means of it. The revenue of the society consists altogether in those goods, and not in the wheel which circulates them” (Smith 1776:385).

Despite Smith’s analysis, the idea of money as the instrument of output circulation has been lost in more recent economic thought, which focuses on the stock dimension of money instead of investigating its flow nature. Moreover, money flows are often considered as if they were money stocks “on the wing”, to use Robertson’s (1922) terminology. “One man’s spending is another man’s receipts. One man can reduce his nominal money balances only by persuading someone else

to increase his” (Friedman 1987:4). This is the famous “hot potato game” described by Tobin (1987:273). In such a framework, therefore, analysis can only focus on how money and non-money goods end up in the individuals’ portfolios. As Ingham observes, “[t]here is no attempt to account for the ‘concept’ of money as a *measure of value* (or *unit of account*)—or even to recognize that this might constitute an intellectual problem” (Ingham 1996:515).

### Endogenous Money and Credit

The measure-of-value conception of money and its related function as a means of payment have been addressed, and explained, by opponents to the exogenous money paradigm. Generally speaking, endogenous money proponents argue that the money creation process is credit-driven and demand-determined, as shown in Figure 2 (Moore 1988, Lavoie 1999, Rochon 1999). This paradigm has usually been traced back to Keynes (1930), and Kaldor (1970), although many other well-known and less well-known twentieth-century economists have to be included within this strand of thought, which is akin to the banking school tradition (Realfonzo 1998, Rochon 2001, Gnos and Rochon 2003). Whereas advocates of money exogeneity focus on exchange to explain money’s nature, value, and functions, adherents to the endogeneity-of-money view focus on production and particularly on the banks’ role in financing it. As Lavoie (1984:774) puts it, “[m]oney is introduced into the economy through the productive activities of the firms, as these activities generate income”. In this framework, it is not possible to separate real from monetary factors (Rogers 1989). Barter systems and monetary models based on pure exchange are to be rejected on

logical and factual grounds: production, investment, and money are linked through the banking system. Hence, banking and production systems always contribute together to the determination of income, output, and employment.

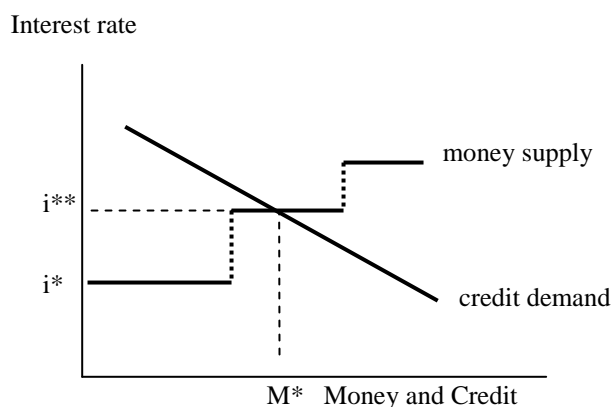


Figure 2. Endogenous Money and Credit

Now, although in this framework the creation of money is essentially tied to bank credit, money and credit—both endogenous here—are separate things. At this juncture, however, the theoretical positions within the endogenous money approach differ, and one can find three separate points of view. First, post-Keynesians argue that “money is bank liabilities, that is, deposits” (Chick 2000: 130), and consider that “the feature which distinguishes money from credit is the general acceptability of deposits, as against the personal quality of credit” (p.131). Secondly, adherents to the monetary circuit approach consider that “money is, and has always been, a debt created *ex nihilo* by bank credit advances that are granted either to permit the generation of real wealth or to acquire existing physical assets” (Parguez and Seccareccia 2000:102). Thirdly, proponents of the theory of money emissions explain that “money is a flow whose instantaneous circulation has a stock of income (or capital) as its object. Banks

create the flow but not its object, which is closely related to production. This is to say that money and credit are not one and the same thing” (Cencini 2001:3). A possible way to integrate these views would begin by analytically separating money (a flow) from bank deposits (stocks): each time that an agent is paid, he is the beneficiary of a flow of money that is instantaneously spent for the purchase of a bank deposit (see Rossi 2001:ch 4). In fact, bank deposits exist as a result of the monetisation by banks of firms’ production costs. Their purchasing power has therefore nothing to do with the general acceptability of the money stuff or with the credibility of the banking system: it depends on the association of money and output that occurs on the factor market when wages are paid out.

In its most radical form, the endogenous money view argues that “the monetary base of the economy [...] expands and contracts endogenously in response to developments elsewhere in the system, and regardless of the nature of the exchange rate regime” (Smithin 1999:1077). However, different authors interpreted the meaning of endogenous money in different ways. Wray (1990) identifies four main views. First, the neo-Chartalist view adopted by the so-called Kansas-City school follows Knapp’s (1924) conception of money as a unit of account sanctioned by the state in the payment of taxes. This school thus explains endogenous money following the taxes-drive-money approach (see also Lerner 1947), arguing that “[m]ost paper money (today, mostly deposits) is privately issued and derives its demand not from a promise of redeemability but rather from state acceptance at pay offices” (Wray 1998:28). Secondly, the followers of Minsky (1957a,

1957b) focus on the role of modern money as a debt used to finance asset positions, and thus explain business cycles by financial factors that may lead to economic crises, although modern institutions might prevent them (see Minsky 1982). According to this view, bank money is endogenous owing to financial innovation that weakens the monetary authority’s role in controlling the money supply, which was originally exogenous and that the evolution of the banking system has in part turned into endogenous (see Chick 1986, Pollin 1991, Palley 1996). Thirdly, several authors led by Davidson (1972) explain that money is endogenous owing to the existence of non-probabilistic uncertainty and risk in all areas of economic activity.

In fact, these scholars are concerned with money as a time device to protect agents against uncertainty and risk in a non-ergodic world, which is not necessarily prone to crises, but which cannot guarantee full-employment output owing to the pervasiveness of uncertainty, and agents’ liquidity preference and hoarding (see Dow & Dow 1989, Bibow 1998, Dalziel 2001). Fourthly, other endogenous-money writers follow Lavoie (1984) and focus on money as a flow, emphasizing the reflux principle as the mechanism by which money is destroyed. These writers stress the fact that money is endogenous on account of the banks’ policy of responding to the demand for loans from creditworthy borrowers (see Moore 1988:chs 2–3, Lavoie 1992:Ch. 4, Rochon 1999:chs 1–2). Despite these, and some other, differences (see Musella 1999:260, Seccareccia 1999:765–6, Rochon & Rossi 2003:xxv–xxxviii), there are two key causal relations with which all proponents of money endogeneity agree: loans make deposits, and deposits create

reserves. So, “the supply of money is a *consequence* of increased loan expenditure, not the *cause* of it” (Kaldor & Trevithick 1981:5). The money stock is the result of a money creation process driven by the agents’ creditworthy demand for bank loans, which, when banks are agreeable, gives rise to bank deposits. The central bank sets the base rate of interest, and can thus influence the whole structure of interest rates.

In this respect, endogenous money writers have been developing their views along two lines of thought, the accommodationist and the structuralist approaches, which, in fact, might be considered as complementing each other, since they address different questions within the endogenous money view rather than providing different answers to the same question (Fontana 2003). “One perspective argues that when banks and other intermediaries hold insufficient reserves, central banks must necessarily accommodate their needs. To act otherwise would threaten the viability of the financial structure, and hence of the overall economy” (Pollin 1991:367). This amounts to arguing that the whole supply of money (including high-powered money, that is, bank reserves or, in modern parlance, settlement balances for interbank transactions) is endogenous, irrespective of whether we are in an asset-based financial system or in an overdraft economy (Lavoie 2005). The monetary authority, however, is able to determine the short-term interest rate at which reserves are made available to banks. In a nutshell, in the money market the central bank acts as price maker and quantity taker. By contrast, the second perspective considers that monetary authorities may decide to curb the expansion of reserves, thus not fully

accommodating the banks’ demand for them. In this case, however, they set off a process of financial innovations, both of an asset and a liability type, that are bound to give rise to reserves endogenously.

Evidence of such innovation induced by a less than fully accommodating central bank includes liability management practices that in the 1970s and 1980s led several banks to offer Eurodollar deposits and certificate of deposits (CDs), on which reserve requirements imposed by the central bank were much less restrictive than those imposed on sight (or demand) deposits. Still more recently, technical changes in privately managed (large-value) payment systems led to the introduction of fully electronic fund transfer systems and e-money products in a number of OECD countries, which reduced the need for central bank money in interbank transactions (OECD 2002, BIS 2003). The result of such innovations is twofold: “(a) a given volume of reserves will support more liability-managed type deposits, and (b) a given volume of demand deposits will support more bank loans to business” (Pollin 1991:375).

Now, as noted by Pollin (1991:375–6), the practice of liability management puts an upward pressure on the structure of interest rates, owing to the fact that banks have to induce their clients to switch from demand deposits to other forms of deposits on which a lower (or zero) reserve requirement exists. To attract clients to these alternative forms of deposits, banks must offer higher interest rates, and this shifts the structure of interest rates upward. In this connection, considering the liquidity preference of all kinds of agents (households, firms, banks, as well as central banks) leads to elaborate a general theory of endogenous money, where the

structure of interest rates may be affected by the agents' changing preferences for holding a part of their wealth in a liquid store of value (that is, in the form of bank deposits) rather than in a less liquid asset (Dow & Dow 1989). In particular, when the liquidity preference of both households and banks rises, as was observed in the 2000s after the burst of the speculative bubble (new-economy assets) in the US and European financial markets, the economic system could run into a liquidity shortage that affects production as well as employment. This situation contributes to, if not provokes, credit crunches and may even lead to a financial crisis, which in this view is thus an endogenous phenomenon. Further, as shown by the Argentinean crisis, "if a rise in liquidity preference takes the form of a preference for foreign over domestic assets, it will itself cause expectations of exchange [rate] depreciation which will then generate further capital outflows" (Dow & Dow 1989:155–6).

The policy implications that derive from an endogenous-money analysis of monetary economies of production and exchange are therefore manifold, and may be summarised briefly as follows. Contrary to the so-called "new consensus view"—which rejects on factual grounds the exogenous-money hypothesis of the quantity theory of money, but which retains the main conclusions of this theory, namely, the long-run neutrality of money and monetary policy (see Fontana and Palacio-Vera 2003)—the endogenous-money view shows the long-run effects on output, hence on employment, of changes in the structure of interest rates as a result of changes in the agents' liquidity preference. During a boom, when liquidity preference is generally falling, a central

bank's hike in the base interest rate, to curb inflation, induces banks to raise their interest rates in order for them to keep interest rate spreads unchanged (on the assumption of a constant mark-up). This upward pressure on the whole structure of interest rates is further reinforced by the central bank's restrictive behaviour on bank reserves, which leads banks to asset- and liability-manage their balance sheets. This slows down investment, therefore aggregate demand, and renders firms more vulnerable to cash-flow problems (an issue that Minsky explained with his financial instability hypothesis). As a result, the banks facing the risk of default by borrowers in financial troubles, as well as the likelihood of capital losses on their own investments, may decide to limit new loans and might even ask for credit reimbursements, thus aggravating the borrowers' situation. All in all, as explained by Dow and Dow (1989:161), "[t]his phenomenon adds additional force to the production cycle, and indeed determines to a considerable extent the level of output and employment at which the peak occurs".

### **Selected References**

- Balasko, Yves and Karl Shell. (1981) "The Overlapping-Generations Model. II: the Case of Pure Exchange with Money", *Journal of Economic Theory*, Volume 24, Number 1, pp. 112–42.
- Bank for International Settlements. (BIS) (2003) *The Role of Central Bank Money in Payment Systems*. Basle: BIS.
- Bibow, Jörg. (1998) "On Keynesian Theories of Liquidity Preference", *The Manchester School*, Volume 66, Number 2, pp. 238–73.
- Bofinger, Peter. (2001) *Monetary Policy: Goals, Institutions, Strategies, and*



- Instruments*. Oxford: Oxford University Press.
- Cannan, Edwin. (1921) "The Meaning of Bank Deposits", *Economica*, Volume 1, Number 1, pp. 28–36.
- Cencini, Alvaro. (1995/1997) *Monetary Theory, National and International*. London and New York: Routledge.
- Cencini, Alvaro. (2001) *Monetary Macroeconomics, A New Approach*. London and New York: Routledge.
- Chick, Victoria. (1986) "The Evolution of the Banking System and the Theory of Saving, Investment and Interest", *Économies et Sociétés* ("Série Monnaie et Production", Number 3), Volume 20, Number 8–9, pp. 111–26.
- Chick, Victoria. (2000) "Money and Effective Demand", in John Smithin (Editor), *What is Money?* London and New York: Routledge, pp. 124–38.
- Clower, Robert W. (1967) "A Reconsideration of the Microfoundations of Monetary Theory", *Western Economic Journal*, Volume 6, Number 1, pp. 1–8.
- Dalziel, Paul. (2001) *Money, Credit and Price Stability*. London and New York: Routledge.
- Davidson, Paul. (1972/1978) *Money and the Real World*. New York: John Wiley and Sons.
- Dow, Alexander C. and Sheila C. Dow. (1989) "Endogenous Money Creation and Idle Balances", in John Pheby (Editor), *New Directions in Post-Keynesian Economics*. Aldershot and Brookfield: Edward Elgar, pp. 147–64.
- Fontana, Giuseppe. (2003) "Post Keynesian Approaches to Endogenous Money: A Time Framework Explanation", *Review of Political Economy*, Volume 15, Number 3, pp. 291–314.
- Fontana, Giuseppe and Alfonso Palacio-Vera. (2003) "Is There an Active Role for Monetary Policy in the Endogenous Money Approach?", *Journal of Economic Issues*, Volume 37, Number 2, pp. 511–17.
- Friedman, Milton. (1987) "Quantity Theory of Money", in John Eatwell, Murray Milgate and Peter Newman (Editors), *The New Palgrave: A Dictionary of Economics*. London and Basingstoke: Macmillan, vol. IV, pp. 3–20.
- Gnos, Claude and Louis-Philippe Rochon. (2003) "Joan Robinson and Keynes: Finance, Relative Prices and the Monetary Circuit", *Review of Political Economy*, Volume 15, Number 4, pp. 483–91.
- Hahn, Frank H. (1973) "On the Foundations of Monetary Theory", in Michael Parkin and Avelino R. Nobay (Editors) *Essays in Modern Economics*. London: Longman, pp. 230–42.
- Handa, Jagdish. (2000) *Monetary Economics*. London and New York: Routledge.
- Hicks, John R. (1967) *Critical Essays in Monetary Theory*. Oxford: Clarendon Press.
- Hume, David. (1752) "Of Money", in Eugene Rotwein (editor), *David Hume: Writings on Economics*. Edinburgh: Thomas Nelson and Sons, 1955, pp. 33–46.
- Ingham, Geoffrey. (1996) "Money is a Social Relation", *Review of Social Economy*, Volume 54, Number 4, pp. 507–29.
- Kaldor, Nicholas. (1970) "The New Monetarism", *Lloyds Bank Review*, Number 97, pp. 1–18.
- Kaldor, Nicholas and James A. Trevithick. (1981) "A Keynesian Perspective on

- Money”, *Lloyds Bank Review*, Number 139, pp. 1–19.
- Keynes, John M. (1930) *A Treatise on Money*, in *The Collected Writings of John Maynard Keynes*. London and Basingstoke: Macmillan, 1971, vols V–VI.
- Knapp, Georg F. (1924) *The State Theory of Money*. Abridged edition translated by H.M. Lucas and J. Bonar. London: Macmillan (first German edition 1905).
- Lavoie, Marc. (1984) “The Endogenous Flow of Credit and the Post Keynesian Theory of Money”, *Journal of Economic Issues*, Volume 18, Number 3, pp.771–97.
- Lavoie, Marc. (1992) *Foundations of Post-Keynesian Economic Analysis*. Aldershot and Brookfield: Edward Elgar.
- Lavoie, Marc. (1999) “The Credit-Led Supply of Deposits and the Demand for Money: Kaldor’s Reflux Mechanism as Previously Endorsed by Joan Robinson”, *Cambridge Journal of Economics*, Volume 23, Number 1, pp. 103–13.
- Lavoie, Marc. (2005) “Lessons from Asset-Based Financial Systems with Zero-Reserve Requirements”, in Giuseppe Fontana and Riccardo Realfonzo (Editors), *The Monetary Theory of Production: Tradition and Perspectives*. Basingstoke: Palgrave Macmillan, pp. 257–68.
- Lerner, Abba P. (1947) “Money as a Creature of the State”, *American Economic Review*, Volume 37, Number 2, pp. 312–17.
- Mankiw, N. Gregory. (2003) *Macroeconomics*. New York: Worth Publishers, 5<sup>th</sup> edition.
- Minsky, Hyman P. (1957a) “Monetary Systems and Accelerator Models”, *American Economic Review*, Volume 47, Number 6, pp. 859–83.
- Minsky, Hyman P. (1957b) “Central Banking and Money Market Changes”, *Quarterly Journal of Economics*, Volume 71, Number 2, pp. 171–87.
- Minsky, Hyman P. (1982) *Can “It” Happen Again? Essays on Instability and Finance*. Armonk: M.E. Sharpe.
- Moore, Basil J. (1988) *Horizontalists and Verticalists: The Macroeconomics of Credit Money*. Cambridge: Cambridge University Press.
- Musella, Marco. (1999) “Endogenous Money and Credit”, in Phillip Anthony O’Hara (Editor), *Encyclopedia of Political Economy*. London and New York: Routledge, Volume 1, 259–61.
- Organization for Economic Co-operation and Development. (OECD) (2002) *The Future of Money*. Paris: OECD.
- Osborne, Dale K. (1992) “Defining Money”, in Peter Newman, Murray Milgate and John Eatwell (editors), *The New Palgrave Dictionary of Money & Finance*. London and Basingstoke: Macmillan, vol. I, pp. 602–6.
- Palley, Thomas I. (1996) “Beyond Endogenous Money: Toward Endogenous Finance”, in Ghislain Deleplace and Edward J. Nell (Editors), *Money in Motion: The Post Keynesian and Circulation Approaches*. London and New York: Macmillan and St. Martin’s Press, pp. 516–31.
- Parguez, Alain and Mario Seccareccia. (2000) “The Credit Theory of Money: the Monetary Circuit Approach”, in John Smithin (Editor), *What is Money?* London and New York: Routledge, pp. 101–23.
- Pollin, Robert (1991) “Two Theories of Money Supply Endogeneity: Some Empirical Evidence”, *Journal of Post Keynesian Economics*, Volume 13, Number 3, pp. 366–96.

- Realfonzo, Riccardo (1998) *Money and Banking: Theory and Debate (1900–1940)*. Cheltenham and Northampton: Edward Elgar.
- Robertson, Dennis H. (1922) *Money*. Cambridge: Cambridge University Press, 1937.
- Rochon, Louis-Philippe. (1999) *Credit, Money and Production: An Alternative Post-Keynesian Approach*. Cheltenham and Northampton: Edward Elgar.
- Rochon, Louis-Philippe. (2001) “Cambridge’s Contribution to Endogenous Money: Robinson and Kahn on Credit and Money”, *Review of Political Economy*, Volume 13, Number 3, pp. 287–307.
- Rochon, Louis-Philippe and Sergio Rossi. (eds) (2003) *Modern Theories of Money: The Nature and Role of Money in Capitalist Economies*. Cheltenham and Northampton: Edward Elgar.
- Rogers, Colin. (1989) *Money, Interest and Capital: A Study in the Foundations of Monetary Theory*. Cambridge: Cambridge University Press.
- Rossi, Sergio. (2001) *Money and Inflation: A New Macroeconomic Analysis*. Cheltenham and Northampton: Edward Elgar.
- Samuelson, Paul A. (1958) “An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money”, *Journal of Political Economy*, Volume 66, Number 6, pp. 467–82.
- Schumpeter, Joseph A. (1954) *History of Economic Analysis*. London: Routledge, 1994.
- Seccareccia, Mario. (1996) “Post Keynesian Fundism and Monetary Circulation”, in Ghislain Deleplace and Edward J. Nell (Editors), *Money in Motion: The Post Keynesian and Circulation Approaches*. London and New York: Macmillan and St. Martin’s Press, pp. 400–16.
- Seccareccia, Mario. (1999) “Money, Credit and Finance: History”, in Phillip Anthony O’Hara (Editor), *Encyclopedia of Political Economy*, London and New York: Routledge, Volume 2, pp. 763–6.
- Smith, Adam. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*. Harmondsworth: Penguin, 1970.
- Smithin, John. (1999) “Fundamental Issues in Monetary Economics”, in Shri B. Dahiya (Editor), *The Current State of Economic Science (Volume 2 Microeconomics, Macroeconomics, Monetary Economics)*. Rohtak: Spellbound, pp. 1075–91.
- Smithin, John. (2000) (Editor) *What is Money?* London and New York: Routledge.
- Starr, Ross M. (1980) “General Equilibrium Approaches to the Study of Monetary Economies: Comments on Recent Developments”, in John H. Kareken and Neil Wallace (Editors), *Models of Monetary Economies*. Minneapolis: Federal Reserve Bank of Minneapolis, pp. 261–3.
- Tobin, James. (1987) “Commercial Banks as Creators of ‘Money’”, in *Essays in Economics (Volume 1 Macroeconomics)*. Cambridge: MIT Press, pp. 272–82.
- Wallace, Neil. (1980) “The Overlapping Generations Model of Fiat Money”, in John H. Kareken and Neil Wallace (Editors), *Models of Monetary Economies*. Minneapolis: Federal Reserve Bank of Minneapolis, pp. 49–82.
- Withers, Hartley. (1909/1930) *The Meaning of Money*. New York: Dutton.

Wray, L. Randall. (1990) *Money and Credit in Capitalist Economies: The Endogenous Money Approach*. Aldershot and Brookfield: Edward Elgar.

Wray, L. Randall. (1998) *Understanding Modern Money: The Key to Full Employment and Price Stability*. Cheltenham and Northampton: Edward Elgar.

*Sergio Rossi*  
*Department of Political Economy*  
*University of Fribourg*  
*Switzerland*  
*sergio.rossi@unifr.ch*

## Financial Regulation and Deregulation

*Stavros D. Mavroudeas,  
Demophanes Papadatos*

### **Introduction**

Whether finance should be regulated or not and - if yes - to what extent and under what forms has been an issue from the very first steps of the capitalist system. In fact this issue emerged even before, when the first forms of pre-capitalist and primitive capitalist activities started slowly to emerge from within the bounds of feudalism.

During capitalism's emergence finance—one of the main channels of pre-capitalist and proto-capitalist activities—was viewed with suspicion because it was one of the main levers for the subversion of the feudal order and the slow but steady predominance of capitalist relations. This suspicion—and the concomitant attempts to check and regulate financial activities—was dressed with the religious distaste of their origin in money-lending, which was equated with usury. This led, in the predominantly Christian West to money lending being dominated by non-Christian outsiders which, in turn, intensified the distaste for it, even while the services of money-lenders were becoming increasingly demanded.

When capitalism was established the issue of regulation of the financial system retained its importance not because finance was an 'intruder'—as in the case of the feudal past—but because of its crucial importance for the operation of the system. This importance stemmed from the fact that capitalism—compared with previous socio-economic systems—is a heavily monetised commodity system. Additionally, money (and its provision, i.e.

finance) operates as capital; i.e. it is crucial for the acquirement of the necessary means for productive activities. More specifically, finance is a crucial prerequisite for almost any productive enterprise whereas its importance in previous systems was significantly lesser. For all these reasons the financial sector has always been economically, socially and politically sensitive and, consequently, arises the necessity for some form of supervision.

Capitalism's financial system has two main pillars: (a) banks and (b) capital markets. Their functions are quite different, although in recent times there is a tendency for each one to acquire functions of the other. The specific reasons for the demand for regulation of these two industries, although basically the same, have also important differences.

For the banking sector, the basic problem has been its reliance on public confidence. Banks hold only relatively small reserves against deposits (the fractional reserve banking system). This increases their potential profitability but, at the same time, makes them vulnerable to a loss of public confidence, which may cause a run on their deposits. The risk of collapse arises because bank liabilities (deposits) are extremely liquid whereas their assets (loans) are rather illiquid. Consequently, a collapse will result in heavy losses for individual depositors but also to productive enterprises whose operation depends upon the smooth provision of finance by the banks and/or have debts to the defaulting banks.

On the other hand, capital markets (stock exchanges, securities) are much more volatile and prone to exaggerated expectations and speculative activities, which can lead to 'bubbles'. When these bubbles burst, this leads to the abrupt drop

of the price of stocks and possibly to default of the enterprises whose stocks are involved in the bubble. In this case, the effect on individual investors (buyers of company stocks) is evident but the effect on the productive system can be even more direct and acute.

Schools of economic theory approach via different ways the problem of financial regulation. For neoclassical theory, it is mainly a technical matter, which concerns individuals (as the economic agents) and does not have any social and political dimension. This can create significant lacunae because regulation requires a role for the state and/or some other institution and neoclassical theory has no theory for the state and only rather dissenting versions of theories for institutions. On the other hand, political economy—and particularly its Marxist variant—by definition and in principle adopt a much wider perspective where social classes and the role of the state and the institutional framework are organically linked with economic relations.

Finally, the historical dimension is critical. Usually, the evolution of the financial system led to the forms of regulation adopted as specific problems were faced. Thus, when banks provided only a small part of monetary instruments the smooth operation of the system was left to them and, in case of problems, the depositor took the full costs of his choices. When the issue of banknotes expanded and threatened financial stability, the state took over the issue of notes. When bank credit and deposits became widespread then the state and its agencies devised mechanisms of regulation (such as reserve or liquid asset requirements). Of particular importance was the regulatory framework established mainly after the 2<sup>nd</sup> World

War—under the influence of Keynesianism—that managed, both nationally and internationally, the operation of the financial system.

The 1973 crisis, the subsequent capitalist restructuring and the predominance of neo-liberal and neo-conservative policies demolished much of the state regulatory mechanisms of the previous era. During the 1970s through to the 1990s throughout most areas of the world the financial system was deregulated to varying degrees: this included the removal of what was thought as obstacles to competition and thus to the efficiency of the financial sector. Deregulation included the abolition of any controls with regard to the setting of private sector interest rates and the exchange rates as also the significant reduction in the use of variable reserve deposits or the exclusive use of banks for monetary policy actions. As a result, the nature of central banking was also transformed and open market operations have become a dominant part of government actions in financial and money markets, in an effort to become ‘independent’ from governments and to exert their influence on the market more discretely than in the past. However, as the removal of financial regulatory rules increased competition at the same time increased interdependence in the context of the financial system as central banks were weakened. This was exacerbated with the wave of internationalisation of capital (or ‘globalisation’) during the end of the 20<sup>th</sup> and the beginning of the 21<sup>st</sup> century - with its concomitant financial deregulation and the overexpansion of the financial sector. However all these led to serious instabilities both nationally and internationally since the link between finance and production became much more

complicated and unstable. A particularly disturbing sign is the superiority for long periods of returns to financial and particularly speculative activities over returns to productive enterprises. The eruption of several regional financial crises (in Asia, Russia, Latin America etc.) led to the reappearance of calls for new forms of regulation. Within the mainstream policy circles two main opposing approaches can be discerned.

The first maintains that the new *modus operandi* is basically sound and it needs only minor rectifications. Among its proposals are making more transparent the financial sectors of the specific economies on crisis, introducing tougher bankruptcy laws (and even rules governing sovereign states' bankruptcies) to eliminate moral hazard, implementing prudential regulation using the 'Core Principles' of the Basle Committee on Banking Supervision, and allowing greater inflows of foreign capital not only to recapitalise shattered banks but also to 'stabilise' the local financial system by integrating foreign interests. Thus, the general framework and self-regulation suffices to face the problems. Only in special cases specific state intervention is needed, such as the U.S. Fed rescue of the hedge fund Long Term Capital Management after the latter was endangered by the Russian financial crisis. A more heterodox approach calls for new mechanisms of regulation, such as a stronger state regulation of financial markets, a 'World Financial Authority' and the levying of a 'Tobin Tax' on speculative activities. Of all these approaches to the problem of globalization, a consensus seemed to emerge that instability was global and systemic, national efforts would not be sufficient to deal with the problem, and there was a need to reconstruct the

global financial architecture. The ensuing debate (see Akuyz 2000) has concentrated on the following areas:

- standards and transparency;
- financial regulation and supervision;
- management of the capital account;
- exchange rate regimes;
- surveillance of national policies;
- provision of international liquidity;
- orderly debt workouts.

Measures under these headings can help to prevent financial crises, and sometimes serve both objectives simultaneously. Reforms in these areas generally imply significant changes in the operating procedures and governance of the Bretton Woods institutions and most notably the IMF. A number of proposals have been made since the Asian crisis in these areas by governments, international organizations, private researchers and market participants. Some of these proposals have been discussed in international institutions such as the IMF, BIS and the newly-established Financial Stability Forum. Many of the proposals considered in these fora have concentrated on marginal reform and incremental change rather than on the big ideas that emerged on the wake of the East-Asian crisis. More specifically, attention has focused on the transparency, and, to a lesser extent, financial regulation and supervision while efforts have been piecemeal or absent in the more important areas addressing systemic instability and its consequences. Proposals for appropriate institutional arrangements at the international level for global regulation of capital flows, timely provision of adequate international liquidity with appropriate conditions, and internationally sanctioned arrangements for orderly debt workouts have not found favour among the powerful.

Some ‘very important proposals’ did not even make to the agenda of the international community as they were presumably found to be too radical to deserve official attention. Among others these include:

- The proposal for the creation of a global mega-agency for financial regulation and supervision of the World Financial Authority with responsibility for setting regulatory standards for all financial enterprises, off-shore as well as on-shore entities;

- The proposal to establish a genuine international lender-of-last resort (ILLR) with discretion to create its own liquidity;

- The proposal to manage the exchange rates of the G3 currencies through arrangements such as target zones.

- The Tobin tax to curb short-term volatility of capital movements and exchange rates.

Although there are technical difficulties in designing effective global mechanisms for the prevention of financial instability and crises, political constraints and conflict of interest, rather than conceptual and technical problems appear to be the main reason why the international community have not been able to achieve even a modest real progress in setting up effective global arrangements for the prevention and management of financial crises.

### **Mainstream and Heterodox Approaches**

For neoclassical theory the need of financial regulation arises because of two concerns. The first one is that of *systemic risk*, i.e. where the collapse of one financial intermediary can lead to the contagious collapse of others and possibly to the collapse of the whole financial system. This collapse in turn might be transmitted to the real economy. Contagious collapses

of financial institutions can arise from two transmission mechanisms. First, the failure of one financial intermediary may result in a loss of confidence in the financial system in general and, thus, lead to a bank run, which might affect even soundly managed financial institutions. The second mechanism is the very high level of interbank dealings—a common feature in modern finance—where the collapse of a large financial institution might create a significant quantity of problematic debts for other intermediaries and, thus, lead them to default as well.

The second concern with regard to financial collapse relates to *consumer protection*. It is argued that economic efficiency is enhanced when a developed financial system exists and that widespread participation in this system diffuses widely its benefits. However, most people who take part in financial transactions have very little knowledge of their functions. In case of collapse it is difficult to adopt the principle of *caveat emptor* (let the buyer beware). A vast number of small savers would be those mainly harmed and that can have serious economic and political implications. The economic implications can be an exodus from the financial system, which would limit the number of participants as consumers and thus negate the prerequisites for consumer sovereignty. The political implications are that no political system can ignore the plight of a great number of its citizens who have given their funds, one way or another, to the collapsed financial institutions.

The development of the mainstream theory of regulation and deregulation is based on neoclassical microeconomic theory (Goodhart 1989:Ch.9). Regulation becomes necessary only because of the existence of market failure. Market failure



in this context arise from a number of reasons, such as elements of monopoly or oligopoly; because confidence in the financial system is a public good which is not adequately priced from the market; because of the lack of perfect knowledge and so on. However, not all mainstream theorists agree to the extent of the market failure, and consequently on the extent and the nature of the necessary regulation to be applied. This is where the mainstream theory is divided in two versions, the orthodox and the heterodox. Arguments in favor of deregulation are formed in the context of the orthodox approach, while arguments in favor of regulation or re-regulation are formed in terms of the heterodox approach to financial regulation.

In particular, according to the orthodox view, the free operation of the market forces in the sector of finance would achieve efficiency. Only in some limited areas, there is a (limited) possibility of market failure. In these limited cases the role of regulatory mechanisms are necessary. Monetarists restrict Financial Regulation to deposit insurance and lending of last resort (see Schwartz 1987). On the other hand, the heterodox approach to regulation forms the theoretical background for the demand for what we might call re-regulation. According to this view, the free operation of the market forces in the sector of finance would not achieve efficiency. The system has inherent, organic causes that create disequilibria and market failures. Thus, regulation is necessary in general and not only in some limited cases. Asymmetric information is a modern version of this view, which has, however, significant Keynesian origins (Stiglitz & Weiss 1981).

Next we will present the essence of the debate between the mainstream advocates

of deregulation versus those who argue in favour of regulation (or re-regulation). The debate for the form of regulation, that is who should carry out the regulation will follow as a by-product of the main controversy for deregulation versus regulation.

### **Deregulation**

Advocates of deregulation argue that regulation can cause problems as well as solve them. The root cause of these problems is to be found in the involvement of government in regulation. As a result over the past thirty-five years there has been a strong demand for freedom from government control in all areas of economic activity including the financial markets. In the U.S. Federal regulators implemented various deregulatory measures in the 1980s, exploiting crevices in the existing statutory framework. They were prompted to do so by banks that sought new powers in an effort to expand their profitability.

At the theoretical level, in the context of neo-classical economic theory these demands have been supported by the development of the 'economic theory of regulation' (see Stigler 1971:3-31), which stresses a number of undesired features of regulation. The theory derives from the idea of agency capture – that the regulatory process will inevitably be captured by producers and used in their interests rather than in the interests of consumers. The main arguments of this theory, according to Howells and Bain (2002:ch.25), can be summarised in four failings of regulation:

Regulation creates moral hazard. This happens as it establishes a safety net, which causes people to act carelessly. The belief that regulatory rules ensure the safety of deposits with all financial

institutions leads people to deposit their money without giving thought to the credentials of the financial institution to which they deposit. This allows dubious organizations to survive. On the other hand, the certainty that they will be rescued from financial collapse causes also financial institutions to take excessive risk and to over-expand their loans.

There is a danger that the regulatory process might be controlled (captured) by the producers, since the regulatory process has a more immediate and intense effect on the relatively small number of producers relative to each one of the consumers. Moreover, there is the additional issue that many of the regulators move into the industry at some point of their careers and may not wish to confront producers, or the opposite might happen and regulators might be ex-producers having certain ties to the industry. These issues create an additional form of moral hazard.

Regulation creates a cost of compliance for the producers. The burden of such costs depends on the degree of competition in the financial sector and on the character of these costs. If producers are able to pass the burden of the costs to consumers, then this is done through higher prices and a lower output. On the other hand, fixed compliance costs will have no impact on marginal costs and, in a fully competitive industry, will not affect output. In addition regulations may also mean benefits for firms, such as improved record keeping, which might be an advantage at a later stage.

The regulatory burden increases the costs of entry into and exit from markets. This results in the preservation of monopolistic positions in the market and stabilizes cartels.

Moreover, from the last two points, it is argued that regulation constrains competition and reduces the efficiency of the financial markets in the allocation of scarce resources to the economy. However, even if we accept that regulation constrains competition, it is not the only factor in financial markets, which does that, and there is no guarantee that a reduction in regulation will lead to increased competition. Thus, the deregulation argument works also in different ways.

It is argued that regulation prevents new firms from entering the financial market, which, if they could enter, the effect would have been a more efficient market, lower prices, and the devolvement of monopoly rents. In addition, regulation prevents mergers and acquisitions and allows inefficient financial institutions to remain in the industry. Deregulation would lead to mergers, resulting in economies of scale, and on improvements in management.

However, even at the purely theoretical level the lower prices which might arise from deregulation have to be compared to the possible costs due to reductions in the stability of the system and thus to increased cost of loss for depositors. Moreover, studies of the merger activity, which took place in the US Banking in the 1980s and 1990s (Rhoades (1992)), have not produced convincing evidence for increased efficiency and stability. On the other hand, if there were significant economies of scale, and mergers occurred to take advantage of them, it would become difficult for new entrants to come into the market, even in the absence of regulation. It therefore seems that none of the criticisms of regulation provide sufficient reason to reject all regulation, though they do point to matters which must be taken into account in decisions as to

how much regulation there should be and what form it should take.

### **Regulation or Re-regulation: The New-Keynesian Approach.**

Liberalised financial markets have been associated with certain excesses/bubbles during the 1980s which were culminating in the global equity market crashes of 1987, and the Savings and Loans crisis in the US. Briefly among the most important of these excesses of the 1980s are: 1. the increased use of credit to purchase assets and finance consumption, resulting in sharp and perhaps unsustainable increases in personal-sector debt, and concomitant lower saving; 2. Asset price inflation and volatility, with asset prices occasionally deviating sharply from 'fundamentals'; 3. inflation and balance of payments difficulties, arising from the slow adjustment of goods markets compared with financial markets; 4. changes in the effectiveness of monetary policy; 5. financial fragility and systemic risk; As a result of several financial crisis episodes, by the early 1990s, the movement towards financial deregulation, at least in the United States had been largely halted (Litan 1992:55-56). As the 1990s opened, bank regulators in the US, were taking a stricter approach to the enforcement of their capital standards. In particular the Congressional sentiment was in favour of the enactment of legislation designed to stiffen further the power of bank regulators, to ensure the removal of weak and insolvent institutions from the financial system in an orderly manner.

Moreover, during *the late 1980s and the 1990s* other additional factors entered the regulatory environment such as globalization and financial innovation and contributed to the tendency to halt the

deregulation process of the 1980s. (By the term financial innovation is meant profit-seeking investment, diffused through financial markets, which aims to overcome imperfections in those markets (Davis 1995:354)

The greatest effect of globalization on financial regulation came as a result of the huge increase in international interbank lending (Levich 1998:17). Increased international interbank lending resulted in an increased possibility for collapses of financial institutions and for the development of bubbles in capital markets, globally, having contagious repercussions in affecting many countries. In particular, as regards the banking sector, the globalization process increased the possibility that the collapse of a financial institution in one country could have contagious effects to financial institutions in other countries.

Regulators were already aware of such possibility because of certain financial crisis episodes in the early 1970s (Franklin National Bank in New York, Herstatt Bank in Germany). At another regulatory level, globalization also caused fears of competitive laxity among national regulatory authorities. Competitive laxity means that a national authority may feel that too strict regulation may leave its own country's financial institutions at a competitive disadvantage compared to financial institutions based in countries with less stringent rules. All these developments led the way for the establishment of international risk-based capital standards. The (first) 'Basle Accord' was finished in 1988 and made applicable to the banks in the participating countries on a gradual basis, first in 1991 and then in 1993. Capital adequacy is central for the 'Basle Accord' because a

bank's capital must be sufficient to absorb losses and to finance the operation of its business. The modern assessment of a bank's capital adequacy is based on a calculation of its risk-asset ratio. This involves a number of steps: 1. a definition of the elements of capital for supervisory purposes; 2. the allocation of weights to different broad categories of asset (e.g. cash, government securities, loans to banks, loans to firms and households); 3. the expression of capital as a percentage of total risk-weighted assets.

In June 1999, the Basle Committee issued a proposal for a new capital adequacy framework that was to replace the 1988 Accord and became known as the (second) New Basle Capital Accord. The New Capital Accord has been designed to improve the extent to which regulatory capital requirements reflect underlying risks and to address specifically the financial innovation that has occurred in recent years. The new capital framework consists of three pillars: 1. minimum capital requirements, developing and expanding on the standardized rules set forth in the 1988 Accord; 2. a supervisory review of an institution's capital adequacy and internal assessment process; and 3. the effective use of market disclosure to strengthen disclosure and encourage safe and sound banking practices.

Many banks, especially in the United States and Japan, have had to raise their capital to meet the new 'Basle' standards. Even for the surviving thrifts of the Savings & Loans financial episode, the Congress adopted a proposal to tighten capital standards *in 1992*. However, the adoption of the 'Basle Accord' standards for banks was not enough to prevent the globalization of capital markets, which was based on banks as the increased volume of

international interbank lending also meant a huge increase in the mobility of capital on a global scale. Thus, the development of bubbles could be developed now on a global scale. Volatility in capital markets could now be more broadly based. In particular, in the new environment where capital could enter and exit many countries' stock markets within hours, speculation and exaggeration in expectations reached new highs. The result came in 1997 as the 'Asian crises', the first financial crisis of globalization. At the level of neoclassical economic theory the development of the literature on asymmetric information, which forms the kernel of the mainstream heterodox approach provided the needed theoretical support to this so-called re-regulation movement (Stiglitz 1985). For Stiglitz financial markets are inherently prone to market failure due to information imperfections (see Stiglitz & Weiss 1981). Financial markets convey information in order to allocate funds. Yet, because information is asymmetrically available and has a cost of acquisition externalities are created and market failures caused. Therefore, government financial regulation is necessary in order to improve capital allocation. Also, within this argument there is a mild preference of banks over stock markets, as the formers' behaviour is less destabilising and focuses on long-run perspectives. In particular, if information is not freely available it is likely to be distributed unevenly between agents (e.g. banks, enterprises). In other words, if we take any transaction involving two agents, it is probable that one will have better information than the other. Such asymmetric information before the transaction leads to the problem of adverse selection. The latter occurs in situations in

which borrowers with risky projects (which are likely to lead to adverse outcomes) most actively seek loans and thus are more likely to be selected by banks. Moreover, asymmetric information after the transaction leads to the problems of moral hazard. Here the bank is vulnerable to the risk (hazard) that borrower will engage in activities (immoral) likely to lead to default.

### **Self-Regulation versus Government Regulation**

Next, the controversy over the form of regulation followed as a by-product of the discussion for regulation versus deregulation. The main issue is whether regulation should be left to the producer (that is to the market) and be self-regulated, or whether there must be government intervention in the market. Arguments in favour of self-regulation have in their background the proposition that market failure is not an inherent feature of the financial market system. Thus, they argue that an endogenous process might arise, without government intervention, which prevents market failure and achieves efficiency.

On the other hand arguments in favour of the regulation of the market, have as their background proposition that market failure is an inherent feature of the financial system and as a result without governmental intervention it is not possible for the market, if let to itself, to achieve efficiency. In particular, the argument for self-regulation is formed on two elements. First, the industry has an incentive to protect its own reputation and thus a cost will be willingly paid to achieve this overcoming a principal market failure argument in favour of government regulation. Second, market participants

understand the needs of the financial industry and they will interfere less with the efficient functioning of the industry. As a result it is argued that a common problem with government regulation, such as the imposition of excessive safety standards, will be overcome.

There is a danger, however, that self-regulation may turn out to be problematic. It must be supported by some government regulation at least to the extent that firms are legally required to join the industry regulatory scheme. Otherwise, an incentive would be created for some firms to act as free riders, hoping to benefit from any increase in reputation of the industry resulting from the behaviour of firms within the regulatory organization without paying the costs of membership. It is argued that self-regulation will only create less moral hazard than public regulation to the extent that it leaves an element of risk for both consumers and producers. The problem is that once an element of risk exists, the true degree of risk has to be assessed to allow a judgment to be made of the risk/return profile of an investment. This should cause no worries to professionals, but the general public faces two types of difficulty. First, it may be time-consuming and costly to acquire the necessary knowledge to assess risk accurately. Second, the ability of non-experts in a field to assess risk is notoriously poor. Because of that, the possibility of consumers assessing risk at all accurately is a very special event. This, holds especially in markets where asymmetric information is one of its main features, where producers have both an incentive and the capacity to mislead consumers. Further, the inability of consumers to assess the true level of risk ensures that inefficient firms will continue

to stay in business despite a *reduction* in moral hazard.

### **Financial Regulation in Marxist and Radical Political Economy**

In Political Economy the issue of regulation and deregulation is analysed from an entirely different standpoint from the mainstream perspective. As it has already been argued, for neoclassical theory state intervention distorts accumulation. There are only certain minimalist economic functions recognized for the state: securing property rights and managing 'market failures'. For the different streams of political economy (Marxist, neo-Ricardian, radical etc.), state and institutions have an organic role in capitalist reproduction; contrary to the neoclassical notion of stateless accumulation. In this part we will focus on the views of Marxist political economy and some representatives of radical political economy.

Hyman Minsky (1977) has argued, similarly, that financial markets are prone to instability and, thus, regulation is necessary. His *Financial Instability Hypothesis* focused particularly on the endogenous nature of financial fragility. Indeed, he argued that it is the operation-as-usual of capitalism and its financial system that lead to destabilisation. That is, a period of stability induces behavioural responses that erode margins of safety, reduce liquidity, raise cash flow commitments relative to income and profits, and raise the price of risky relative to safe assets. During the upswing of the business cycle the demand for new investment causes an excess demand for finance. This increases interest rates which, in turn, increase debt finance and cause a shift from long-term to short-term finance

and from hedge to speculative or Ponzi finance and, finally, a reduction in the safety margins for financial institutions. Additional interest rate increases can create a refinancing crisis and cause firm defaults. Minsky (1963) maintained that forms of regulation (such as the central bank's role as lender of last resort and the stabilizing role of a large government) are necessary in order to mitigate the risks of financial and banking crises or to attenuate their consequences.

Before proceeding further it is necessary to clarify that for Marxist political economy the notion of state covers also a number of other institutions that might appear as independent. It is well known Marx's dictum that, in the end, civil society institutions are nothing more than parts of the political society. That means that independent, voluntary and non-governmental institutions and associations are, in essence, part of the capitalist state mechanism.

For marxist political economy the state has an integral role in capital accumulation and, among others, in the operation of the financial system. The capitalist state has two primary functions with respect to capital accumulation. First, it has to secure the wage labour system; that is to guarantee the existence and the reproduction of a class of free waged workers and the smooth operation of the wage relation. Second, it must secure the monetary system. Generalised commodification under capitalism necessitates the universal acceptability of the monetary means and, therefore, some degree of state or para-statal regulation of the monetary system (De Brunhoff 1976).

In fulfilling these two primary functions the capitalist state has to operate as a *collective capitalist*. That is as an agent

overseeing and guaranteeing the general interests of the capitalist system and, therefore, rising above the various short-term interests of individual capitals. Thus, the capitalist state has to mediate between the often-contradictory interests of different individual capitals.

The particular functions of the state regarding the regulation of the financial system stem from one of its primary functions regarding capital accumulation: the state must guarantee the proper functioning of the financial system over petty individual capitalist interests and excesses. That means it has to secure the credibility of the monetary unit and the smooth provision of financial means. These tasks have certain very tricky aspects, which cannot be tackled without state regulation.

Capital exists in three basic forms: productive, money and commodity capital. Each one has its own specificities and modalities. Of course, the fundamental form is productive capital under the auspices of which surplus value is created.

The financial system is the field of operation of money capital. Money in capitalism has three basic functions to fulfil. Apart from being a standard of measure and a unit of payment – as the orthodox theory suggests – it operates as capital as well. That means that it is directly linked to productive activities by providing them the necessary monetary means. Money as capital is related to the form of interest-bearing capital. Specifically, interest-bearing capital is money capital traded as a commodity and commanding the payment of interest. This form of capital has a dual character. On the one hand it is immediately related to the sphere of real accumulation for the payment of interest. For Marx, the interest

is part of the surplus-value extracted exclusively in the sphere of real accumulation. On the other hand – as it is immediately related to the form of credit money – it has certain degrees of freedom towards the sphere of real accumulation as the rate of interest, which determines interest-bearing capital is formed outside of it (Itoh & Lapavistas 1999:ch.3).

In the context of Marx's circuit of capital the equilibration between real saving and investment is determined by factors related to the sphere of production and that is why it is determined by technological and social parameters, related to forms of capital at the level of production. On the contrary, the balance of hoarding – that is decisions for hoarding and dishoarding taken independently by sections of the capitalist class – determine the short-run rate of interest related to the form of interest-bearing capital, being a monetary phenomenon in the sphere of exchange. It is stressed that money hoarding rather than saving provides the foundation and wherewithal of interest-bearing capital (loanable money capital). Saving, i.e. income of firms and workers that is not consumed is obviously a broader concept than hoarding. The fact that saving takes a monetary form is important (since that imparts additional uncertainty to its transformation into commodities) but saving also has a use-value dimension: it is real output not consumed. Hoarding has no dimension other than the monetary one: it is money that has temporarily lain idle.

Consequently, if hoarding provides the foundation of interest-bearing (loanable) capital and the credit system, it cannot be at all assumed that the operations of lending and borrowing equalize saving and investment. The fact that the flows of interest-bearing (loanable) capital are

mediated by the credit system (with its own markets, traders, prices, and speculative activities), far from leading to equilibrium, might contribute to disequilibria between saving and investment. As this form of capital is primarily formed outside the circuit of capital and enters and exits the circuit, its price (the rate of interest) which determines the balance of hoarding, is primarily determined outside the circuit and does not equate saving with investment. This distinction forms the basis for a demand for financial regulation in the context of capitalist system aiming to the restriction and control of the autonomous character of the form of interest-bearing capital. Financial regulation will mitigate the problems caused by the relative autonomy of this form of capital to the financial system and even further to real accumulation.

On the basis of these general considerations, Marxist political economy studies the specific historical evolution of waves of regulation and deregulation. This is done mainly from within the premises of *periodization theory* (i.e. the distinction of different historical stages of capitalist development). Thus, for the monopoly capital approaches, the stage of *laissez-faire capitalism* (who was characterized by minimal state regulation) was followed by *monopoly* and *state-monopoly capitalism* (in which state economic intervention and regulations assume a major role) from the end of the 19<sup>th</sup> century till today (see Fine & Harris 1979). In monopoly capitalism the importance of the financial system was increased because—due to the centralization of capital, which was also reflected in the predominance of the joint-stock company—productive activity required increased monetary means. Much of the regulation of the financial system,

during that stage, was performed through mechanisms relating different individual capitals (such as *finance capital*, i.e. the fusion of money and productive capital under the auspices of the former and in the form of a bank supporting and dominating a group of enterprises). The state supported this type of operation but without having a direct role.

However, in the subsequent stage of state-monopoly capitalism, the state has a more direct role in the control and regulation of the financial system as the latter becomes of paramount importance for the reproduction of capitalism. Finally, there is much controversy over what is happening after the 1973 crisis, which signified, possibly, the end of the previous stage and the beginning of a new one. For other views there is a general and unequivocal retreat of financial regulation whereas other views stress that the state and its mechanisms have retained their crucial functions and acquired even more while pushing for deregulation or self-regulation of the less important functions.

A special school, the Regulation approach (RA), has tried to answer these questions. This school - stemming from the Marxist political economy and having moved since to radical political economy - has a particular, not only in name, relevance to the issue of financial regulation. The RA argues—in more institutionalist than Marxist terms—that the capitalist economy cannot function properly without the existence of a set of institutional arrangements that ‘regulate’ it; that is, they secure the functionality of its fundamental economic structures. In its original versions, the fundamental sphere was production whereas circulation and the financial system had a significant but secondary role. The regulation of the



financial sector was necessary in order to secure the historically specific requirements of capitalist reproduction.

Thus, in the era of Fordism, the socialisation of finance (through social insurance funds) and its regulation by the state secured against the temporary loss of direct wages and to enable wage-earners to purchase goods whose exchange-value was relatively large. Then, it argued that capitalism—after the 1970s—surpassed Fordism and has entered a new era, which was previously branded as post-Fordism – a very popular but also non-accurately defined concept.

Later, a new answer has been offered by some important Regulationists (Aglietta 2000, Aglietta & Bretton 2000, Boyer 2000). In a nutshell, they argue that possibly, at the end of the 20<sup>th</sup> century, a new finance-led growth regime has been established. Financial deregulation has enhanced the market for corporate control. This obliges firms to retain a part of their profits in order to pay dividends and to boost the price of their equities (that is to conform with shareholder value requirements). The result is that fewer funds are left for productive investment and this, in turn, lowers the growth rate. The stability of this equity-based regime depends crucially on monetary policy and its ability to control financial bubbles. Their diagnosis is that the currently present institutional framework is not sufficient to cope properly with these problems.

This analysis has significant merits and, in some sense, it replicates crucial aspects of the classical Marxist analysis. A typical example is the explanation of the divergence between the rate of return on capital (which is determined in production) and the rate of interest (which is determined by the financial system).

However, the major deficiency of this new Regulationist version is that it—once again—changes the architecture of its theory: instead of production it is now finance that becomes the pivot centre of the system. This new structure changes the balance between productive and unproductive activities via the dominance of finance on industry. This is a critical and unwarranted thesis, which invalidates whatever other merits their analysis has. First, it is highly debatable whether the ‘financialisation’ of the economy during recent decades is a permanent – and not a short-term conjectural – characteristic of contemporary capitalism. Second, it has great similarities with Hilferding’s (1910) thesis on *finance capital*, which provides the basis for most definitions of the monopoly capital stage. Where it appears to differ is that possibly emphasis is placed upon capital markets rather than banks. Notwithstanding, this formulation might have the same deficiencies with Hilferding’s version, i.e. generalise unwarrantedly a limited historically and geographical phenomenon.

### **Selected References**

- Akyuz, Y. (2000) “The Debate on the International Financial Architecture: Reforming the Reformers”, UNCTAD, *Discussion Paper* No.148.
- Aglietta, M. (2000) “Shareholder Value and Corporate Governance: Some Tricky Questions”, *Economy and Society*, Volume 29, Number 1, pp. 146-159.
- Aglietta, M. & Breton, R. (2001) “Financial Systems, Corporate Control and Capital Accumulation”, *Economy and Society*, Volume 3, Number 4, pp. 433–466.
- Boyer, R. (2000) “Is a Finance-Led Growth Regime a Viable Alternative to

- Fordism? A Preliminary Analysis”, *Economy and Society*, Volume 29, Number 1, pp. 111-145.
- Davis, E.P. (1995) *Debt, Financial Fragility and Systemic Risk*. London: Oxford University Press.
- De Brunhoff, S. (1976) *The State, Capital and Economic Policy*. London: Pluto.
- Fine, B. & Harris, L. (1979) *Rereading Capital*, New York: Columbia University Press.
- Goodhart, C.A.E. (1989) *Money, Information and Uncertainty*. Second Edition, London: Macmillan.
- Hilferding, R. (1910) *Finance Capital*. London: Routledge (1981).
- Howells, P. & Bain, K. (2002) *The Economics of Money, Banking and Finance*. London: Prentice Hall.
- Itoh, M. & Lapavitsas, C. (1999) *Political Economy of Money and Finance*. New York: Macmillan.
- Litan, R.E. (1992) “Financial deregulation and Re-regulation”, in P. Newman, M. Milgate and J. Eatwell (Editor), *The New Palgrave Dictionary of Money, Banking and Finance*. London: Macmillan.
- Levich, R. (1998) *International Financial Markets: Prices and Policies*. London: McGraw-Hill.
- Minsky, H. (1963) “Can “It” Happen Again?” in D.Carson (Editor), *Banking and Monetary Studies*. Homewood, Illinois: Richard D. Irwin.
- Minsky, H. (1977) “The Financial Instability Hypothesis: An Interpretation of Keynes and an Alternative to “Standard” Theory”, *Nebraska Journal of Economics and Business*, Volume 5, pp. 45-67.
- Rhoades, S. (1992) “Banking acquisitions”, in P.Newman, M.Milgate and J.Eatwell (Editors), *The New Palgrave Dictionary of Money, Banking and Finance*. London: Macmillan.
- Stigler, G.J. (1971) “The Theory of Economic Regulation”, *Bell Journal of Economics and Management Science*, Volume 2, Number 1, pp. 3-31.
- Stiglitz, J.E. & Weiss, A. (1981) “Credit Rationing in Markets with Imperfect Information”, *American Economic Review*, Volume 71, Number 3, pp. 393-410.
- Stiglitz, J.E. (1985) “Credit Markets and the Control of Capital”, *Journal of Money and Banking*, Volume 17:133-52.
- Schwartz, A.J. (1987) “The Lender of Last Resort and the Federal Safety Net”, *Journal of Financial Services Research*, Volume 1, pp. 77-111.

Stavros D. Mavroudeas,  
Demophanes Papadatos  
Department of Political Economy  
University of Macedonia  
Thessaloniki, Greece  
smavro@uom.gr

## Fiscal Federalism

*Duc Hong Vo*

### **Introduction**

Nation states are traditionally classified in terms of their federal or unitary political system, depending on their written or unwritten constitutions. Within unitary countries: political and fiscal authority is generally centralised in the hands of the national government; the powers and responsibilities of subnational governments are usually limited; and subnational arrangements may be ad hoc. In contrast, the powers and responsibilities of the various levels of government in federal countries are generally constitutionally guaranteed and clearly allocated among different tiers of governments. The extent to which fiscal responsibilities are shared varies from federation to federation. However, to some extent, subnational governments are generally expected to exert more autonomy when making policy decisions in their respective jurisdiction than in the case for unitary countries.

A typical federal constitution provides for the assignment of responsibilities for service provision across all tiers of government, with subnational governments having some degree of autonomy over their own spending programs. Similarly, there is an assignment of taxing powers with subnational governments having some degree of autonomy over their tax rates and bases. The assignment of service provision responsibilities and tax powers are fundamental to fiscal federalism. In addition, there are two other significant issues. The first issue is intergovernmental financial transfers, which usually take place from the national, or higher levels of government, to the lower levels of

government. The second issue is borrowings by subnational governments. These last two issues are particularly important when there is a mismatch between service provision responsibilities and revenue assignment of subnational regions. In total, expenditure assignment, assignment of revenue, intergovernmental financial transfer, and subnational borrowing represent the four pillars of fiscal federalism.

The early literature on fiscal federalism emphasised intergovernmental transfers. The more recent literature has focused on the policy relevance of changing the fiscal constitution by varying the degree of fiscal authority assigned to subnational governments. In the long term, when a relationship among the four pillars changes, particularly the assignment of spending and revenue, there must be a change in the “degree” of fiscal autonomy among different tiers of government. In practice, federal states are usually more decentralised than unitary states. This can be partly explained by the fact that federal countries have a tendency to assign greater powers and responsibilities to lower levels of governments. However, this does not guarantee that federalism is a necessary condition for decentralisation.

### **Early Literature on Fiscal Federalism**

Tiebout’s classic article “A pure theory of local expenditures” was published in 1956. Since then, over the next half a century, the field of fiscal federalism has developed substantially and contributed to a large body of literature. Subsequent seminal studies by Musgrave (1959) and Oates (1972) laid the foundation for significant discussions of fiscal federalism.

Tiebout (1956) introduced the notion of local public expenditures to demonstrate

that, in a fiscally decentralised country, perfect mobility of citizens among localities will result in competition among localities. When tax and rate payers vote with their feet, local governments are forced to compete to attract tax and rate payers by supplying them with local public services at minimal costs (productive efficiency) and in the quantity, quality and range that best matches the demands of local communities. As such, Tiebout regarded an appropriately devolved fiscal federalism as a mechanism for overcoming the free rider problem for local public goods. Three years later, Musgrave (1959) laid the general foundations for modern public finance theory, stressing that the best allocation of scarce resources will be achieved whenever preferences and tastes of local citizens have been met. Oates (1972) subsequently argued that there should be variations in the provision of public goods and services since inhabitants in different regions have different tastes in their consumption patterns. From this perspective, subnational governments will be more responsive to citizen demand in comparison to the national government, which tends to provide the same bundle of goods and services across regions without regard to regional variations in tastes and preferences. Also, if the national government is the only provider of public goods and services for the community, production efficiency will not be improved due to the non-existence of competition, whereas subnational governments have to face fierce competition from the neighbouring governments.

Oates (1972:35) formalised this issue by developing and proving the *first decentralisation theorem*: “For a public good—the consumption of which is defined over geographical subsets of the

total population, and for which the costs of providing each level of output of the good in each jurisdiction are the same for the central or the respective local government—it will always be more efficient (or at least as efficient) for local governments to provide the Pareto-efficient levels of output for their respective jurisdictions than for the central government to provide *any* specified and uniform level of output across all jurisdictions”.

Subsequent to Oates’ theorem, discussions on fiscal federalism have centred on four main areas: the assignment of responsibility for service provision; revenue assignment; intergovernmental financial transfers; and subnational borrowing. Each of these four pillars has its own implications for efficiency and equity, which should be taken into consideration.

### **Responsibility for Service Provision**

*The principle of subsidiarity* suggests that economic performances should become more productive, efficient and effective if goods and services are provided by the lowest level of government possible. Foreign policy, defence, immigration, and international trade can be best formulated and implemented by the national government. This is done in order to efficiently supply national public goods and to take advantage of economies of scale and scope. These increases in production efficiency are generated as the quantity of public goods and services being produced and provided rises (scale); and as potential cost savings flow from joint production (even if services provided are not directly related to each other) (scope). In addition, the national government is the sole agent willing to increase levels of

spending, if required, to bring full benefits as possible for the entire economy through economic spill-over or positive externality. Services provided by the national government are in accordance with the law of subsidiarity when demand is at a constant level across various regions or states. However, when demand varies from location to location, national provision to a common standard leads to inefficient under-provision, in some areas, and inefficient over-provision, in other areas.

On the other hand, subnational governments are able to carry out some important tasks for their communities. These include law, order and public safety, education, health, as well as very local issues such as street lighting, local sewerage, garbage collection, and footpaths. They can provide goods and services based on their size of jurisdiction, and in accordance with local tastes and preferences (Shah 2004). If the size of jurisdiction is considered, *the principle of benefit matching* will be achieved. This principle suggests that benefits received should match with costs borne of services provided. This sheds light on local provisions of goods and services to which local citizens who receive benefits also bear costs. In addition, the cost of publicly provided goods and services can be recovered, at least at the margin, so that pressures from lack of financial resources decrease.

On the basis of the law of subsidiarity, it is typically efficient for subnational governments to account for a significant proportion of fiscal activity across the nation. The larger the portion of non-national public spending cake, the greater the fiscal importance of subnational governments.

In the case of Canada, a federation, the spending from the national government dominates in some fields with uniquely national interests such as defence, international relations, social welfare and health, etc. Provincial governments in Canada also have significant discretion when spending for education, law, order, and communications. With almost 5,000 local governments, this level of governments is mainly responsible for spending on recreational and cultural areas, local roads and parks.

### **Assignment of Taxing Powers**

Agreement on the distribution of taxing powers is difficult since the players (national and subnational governments) approach their respective powers from two different perspectives. While the national government requires some tax sources for economic stabilisation and income redistribution, subnational governments focus on taxing powers to generate stable revenue to provide local public goods and services which are fundamental to community welfare. For example, in Australia, the national government fully controls the proceeds from income tax, excise tax, import duties—three main important sources of the national tax system. The dominance of the national government leaves state governments with the powers to collect revenue from stamp duty, payroll tax and some other relatively small-revenue-generating taxes. Taxing at the national level provides a foundation to avoid unnecessary competition among states in the country, to efficiently control tax bases and to reduce compliance costs. If devolution of taxing powers is extensive, the financial gap between spending responsibilities and taxing powers of subnational governments will be

minimised or eliminated, leading to their fiscal autonomy.

Fiscal autonomy of subnational governments implies that, to some extent, they should be able to arrange their own sourced revenue by exercising their taxing powers to cover costs occurring in the provision of public goods and services. It should be noted, however, that even in the absence of fiscal transfers (“grants”), subnational governments will not enjoy full fiscal autonomy if they receive taxes and/or revenue shares directly from the national government and all tax bases and rates are centrally determined (McLure, 2001). This case may be illustrated by the example of Vietnam, where proceeds from the Value Added Tax (VAT) are allocated to the provinces, however the tax base is defined and the rate is set by the central government. A necessary condition for a significant level of fiscal autonomy is that subnational governments have the discretion to set tax rates and/or bases. In this case, they can adjust their revenue by varying the rates and/ or the bases in response to fiscal demand for publicly provided goods and services.

There are few countries where subnational governments can independently set tax rates and bases. A popular practice is named “piggybacking”. This is a common practice to grant more fiscal autonomy to subnational governments than permitted through intergovernmental grant. Piggybacking allows subnational governments to set the tax rate as a surcharge on the taxes collected by the national government. Also, tax rates set by subnational governments are only allowed to vary within the band centrally determined which obviously acts as a bounded constraint on their fiscal autonomy, resulting in only partial fiscal

autonomy. However, piggybacking can be difficult to implement in developing countries, mainly because the marginal tax rates of the national government are usually set very high, so that there is no room for an additional “surcharge” on the same tax bases.

### **Intergovernmental Fiscal Transfers**

In the event of a long-period mismatch between subnational governments’ spending responsibility and revenue capacity, a “vertical fiscal imbalance” emerges. This imbalance implies a mismatch between own sourced revenue (excluding grants from other levels of governments) and own purpose spending (excluding grants to other levels of governments) for a particular level of government (Collins 2001). Vertical fiscal imbalance is the gap in terms of taxing and spending between the national government and subnational governments.

For example, consider the hypothetical case where a central government raises 80% of total government revenue and spends, say, only 50% on its own revenue. Vertical fiscal imbalance arises because the central government has a surplus of 30% of total government revenue, and the subnational governments have a similar equivalent deficit. As a result, vertical fiscal imbalance must be managed by governments through intergovernmental fiscal grants and advances. If subnational governments acquire adequately fiscal autonomy, vertical fiscal imbalance will be minimised. This happens at least for the aggregate of subnational governments. It is still possible, of course, for individual subnational governments to run deficits or surpluses, even when the vertical fiscal imbalance to be eliminated between the national government and the aggregate of

the subnational governments. In this sense, fiscal balance is only achieved when the national government's matching grants are used to fund subnational governments' deficit.

If the national government focuses exclusively on funding vertical fiscal issues through grants to subnational governments, the incentive for the subnational governments to increase their taxing powers, and to manage public spending efficiently will be reduced (Ahmad & Craig 1997). Nevertheless, fiscal transfers can be seen as a largely unavoidable tendency in both federal and unitary countries because subnational governments are not homogenous entities – some have relatively high revenue raising capacity while others have relatively low revenue raising power. Intergovernmental transfers may also serve other secondary purpose: they may also help the national government achieve its objectives of macroeconomic management. There are a wide range of aims that the national government desires to achieve when designing the system of fiscal transfers. Such transfers may also be used to achieve fiscal equity across vertical and horizontal dimensions.

Boadway (2003) recognised that fiscal equity is matched for unitary countries when their social preferences are the same across regions. In contrast, the notion of fiscal equity in federal countries is more or less ambiguous since national and subnational governments may use a mix of policies to achieve redistribution. Every citizen is exposed to at least two distinct sets of fiscal distribution – one from the national government and another from subnational governments. When there are differences of taxing capability, and, as a result, differences in provision of services

to the given standard among different regions, horizontal fiscal imbalance arises. The imbalance of services provisions across regions causes the so-called horizontal fiscal inequity. This phenomenon can be caused by differences in local wealth, natural resources, per capita tax capacity, and differences in spending obligations and in the cost of providing public services (Martinez-Vazquez 2001).

The criteria of equity and efficiency in fiscal transfers usually go together. Efforts by a national government to fill in the gap among different regions and to achieve fiscal equity may cause inefficiency because resources can be allocated into regions where they cannot be best used. As a result, the balance to meet both criteria of fiscal equity and efficiency is a complex policy issue. Moreover, subnational governments have no incentive to increase the level of services provided to their local inhabitants if they are aware that citizens in the nearby regions also benefit from these services without bearing the full cost. In such a situation, fiscal transfers can encourage subnational governments to increase the quantity of services provided which may create a positive externality for other regions. From an efficiency perspective, the economic benefits to the economy as the whole will increase.

The main tool for achieving equity through intergovernmental fiscal transfers is the mechanism “fiscal equalisation”. Fiscal equalisation is a process which reduces fiscal disparities among subnational regions through financial transfers. There are two types of equalisation: vertical and horizontal. When fiscal transfers are arranged by a higher to lower levels of government to offset vertical fiscal imbalances, vertical

equalisation is achieved if each level of government can balance its budget in aggregate. Horizontal equalisation is achieved when governments at the same level, (for example, state government) have the same capacity to treat like individuals the same by making transfers (usually indirectly) from the “rich” and “poor” subnational governments. For example, identical single supporting parents in both jurisdictions would have access to the same educational opportunities for their child, health services for themselves and for the family and level of police services. In practice, vertical and horizontal fiscal equalisations are usually achieved concurrently. This happens when the national government gives grants to subnational governments (vertical fiscal equalisation), but gives higher per capita grants to poor regions than to rich areas (horizontal fiscal equalisation). Consequently, a system of equalisation payments is basically used to achieve two main objectives. Firstly, it seeks to minimise the vertical fiscal gap between the national and subnational levels of government. And, secondly, it seeks to equalise the horizontal fiscal gap among regions in the federation. Vertical fiscal imbalance is the product of the national government with its intention of introducing efficiency and equity for different regions in the federation through grants from national fiscal resources.

Australia, Canada, and Germany have employed equalisation arrangements substantially to fill in these vertical and horizontal gaps. The US, in contrast, is more focused on vertical imbalance, leaving significant gaps in horizontal equity.

From the perspective of a fiscal autonomy, the principle issue is whether

fiscal transfers to achieve vertical and horizontal equalisation are conditional or unconditional. That is, the extent of discretion the subnational governments enjoy concerning the expenditure of these transfers. While the issue of when to, and how to, equalise transfers is important in the context of horizontal equity, it is generally a second order issue in the context of vertical equity. In regard to vertical equity, the first order question is how to assign responsibilities for services provisions and revenue raising powers in a manner that eliminates vertical fiscal imbalance between the national government and the aggregate of all subnational governments. Even in the presence of vertical fiscal imbalance, the primary is not the value of the intergovernmental transfers required to achieve the vertical fiscal balance, but the proportion in which conditional and unconditional transfers are made to balance vertical fiscal imbalance.

### **Responsibility for Subnational Borrowing**

The autonomy of subnational borrowing is often restricted due to major external events, such as debt crises in Latin America in the 1980s. A concern here that subnational borrowings are not supervised and controlled effectively by the national government. As a consequence, borrowing from subnational governments has been regarded as more risky than national borrowing. However, there are grounds for subnational governments to use debts as a fiscal tool. Firstly, in the circumstances of limited fiscal resources even after intergovernmental fiscal transfers are arranged, subnational governments need capital to cover expenditure for investment purposes. In this sense, borrowing can only



be used for investment into the region. A typical example is borrowings to build toll roads in order to improve local infrastructure of the regions. It is suggested that investment spending cannot be covered by an increase in local recurrent taxes because it will violate the principle of intergenerational equality – the current generation solely bears the costs for future generation's benefits. Secondly, subnational governments' borrowings also reflect the effective link between their autonomy of spending and taxing assignments. This will reveal the accountability of subnational governments to their local inhabitants: markets will signal any poor performance of subnational governments through bad credit ratings which will increase interest rates on public loans or simply blocking access to capital markets. Consequently, the issue is not whether or not subnational government's borrowing is allowed, but the extent of borrowing and the role of the national government in managing and supervising the debt made by subnational governments.

For subnational borrowing, the equity and efficiency implications are quite clear. For the first implication, equity, if vertical fiscal imbalance is not financed by fiscal transfers, it must be financed by borrowing. The principle of equity holds when costs are borne across generations. In addition, for the hard budget constraint, debt made by subnational governments is not guaranteed by the national government. In this case, subnational governments will be fully aware of where to place their money so that benefits can be maximised. This shows that the implication of efficiency holds since money will be put in the most efficient place. In conclusion, subnational borrowings are satisfied on

both efficiency and equity grounds, and, as a result, they should be encouraged.

### **Fiscal Decentralisation and Government**

A nation's degree of fiscal decentralisation, which is now considered the most fundamental issue in fiscal federalism, depends primarily on the level of the assignment of fiscal powers and responsibilities. The issue of fiscal decentralisation has two broad dimensions. First, there is the fiscal autonomy of subnational governments; and, second, their fiscal importance. In this regard, fiscal autonomy of subnational governments mainly deals with the assignment of taxing powers, including supplementary tools such as intergovernmental fiscal transfers, subnational borrowing and the assignment of spending responsibility. Fiscal importance is directly connected with the level of fiscal responsibility that subnational governments enjoy relative to the level of all government expenditures. For example, if the aggregate of subnational governments can generate revenue to meet 80% their expenditure responsibilities, then 0.8 represents for the fiscal autonomy of subnational governments in this example. In addition, if the expenditure of all subnational governments is 50% of all public expenditure, its fiscal importance is 0.5, a ratio of subnational expenditure and total government expenditure.

### **Decentralisation and the Size of Government**

One important argument in support a high degree of fiscal decentralisation is that subnational governments should be involved in providing local services to their region to achieve efficient costs and to

meet local preferences and tastes. If this is the case, it is then expected that the sizes of subnational and national governments will change. The concern is that, at the end, the size of the government as the whole increases or decreases.

Three main arguments support the view of fiscal decentralisation and smaller government. Firstly, the government is seen as a monolithic “Leviathan” always seeking to maximise revenue from taxes by using its power to exploit tax bases and rates (Brennan and Buchanan 1980). Fiscal decentralisation, constitutionally guaranteed, limits governments’ capacity to over-tax. Under the pressure of re-election, governments are required to increase economic efficiency in providing services to save costs. Secondly, while national governments typically apply the strategy of “one-size-fits-all” in providing public services, subnational governments will take local tastes and preferences into consideration so that a better design of public services which can meet local needs can be easily achieved (Oates 1972). This means that resources are efficiently allocated in accordance to local tastes and preferences.

Therefore, fiscal decentralisation is expected to lead to a greater degree of efficiency and a smaller public sector. Thirdly, due to the presence of competition among subnational governments, in terms of local participation, voters at each locality would show their preferences by the votes they were using to elect local officials into power. If local citizens are not satisfied with subnational governments, they may decide to vote with their feet. Local political participation is expected to be at a high level in more fiscally decentralised society and, then, it is able to control the growth of government at local

levels. As a result, the size of the government declines.

Prud’homme (1995) challenged the view of efficiency gains when services are to be provided by subnational governments because he argued that this view focuses mainly on the demand side. As such, subnational governments’ capacity to supply services at sufficient amount is not fully investigated. Also, in developing countries, the basic needs of goods and services are considered essential and important. As a consequence, the combination of public goods and services provided which reflects local preferences and tastes are far from expected. Also, the view on high level of local participation contributing to a decreased size of governments is arguable because this may not be valid in developing countries where local elections are embodied the relationship among tribal and party’s loyalty only. The winners are selected based on the above relationship rather than who can satisfy the needs, wants, tastes and preferences of local inhabitants.

### **Fiscal Decentralisation and Corruption**

With more economic competition for public goods and services provided, there is a force that acts to make subnational governments more efficient, accountable and transparent. This is because local people can easily link local revenue and spending and decide their votes. As a result, corruption should be minimised (Breton 1996). Also, with clear fiscal arrangements among levels of governments, subnational governments should be able to do their job better. Moreover, it is argued that corruption at subnational levels will not be as serious as the national level because the powers of local officials are limited. This implies that

the level of corruption at the local level should be smaller than for the national level.

However, with a higher degree of fiscal decentralisation, the size of subnational governments is widened. As a consequence, many more people become involved with subnational governments. This could lead to a greater incidence of corruption. However, this is likely to be offset by reduced volume of corruption at the national government. The key issue is whether this offset will lead to an increase or a reduction in the overall level of public sector corruption. Furthermore, in the case of developing and transition countries, national government's officials experience good environment of working with high pay, promotion and further training; the government's officials at local levels do not. In such an environment, there may be a higher level of corruption at subnational governments. In addition, within the small administrative region, close relationship becomes an important factor for corruption.

### **Fiscal Decentralisation and Economic Growth**

Fiscal decentralisation is expected to impact favourably on economic growth. There are four main channels in which fiscal decentralisation can contribute to economic growth. The general regime for fiscal decentralisation to be translated into economic growth, controlling all other factors, can be explained as follows.

The first channel is a more efficient allocation of resources. This tends to ensure that local tastes and preferences will be met with minimum costs. Subnational governments are in the better position, compared with the national government, to understand local tastes and preferences. As

a result, costs of provision of public goods and services are minimised. The second channel is competition among subnational governments. The pressure of being re-elected puts local officials to work in a more economic efficiency manner. This dynamic efficiency is attained by minimising wastage and this will be translated into economic growth. Smaller size of governments contributing to economic growth is the third channel. It is argued that smaller government may result in below optimum size relative to the needs of education, health, infrastructure, and etc. However, this argument, in principle, need not be accepted since services can be jointly provided across local governments. A smaller size of government may minimise administrative costs due to a reduction in wastage. As a consequence, economic growth is enhanced through an increase in efficiency. The last channel for fiscal decentralisation to be translated into economic growth is from lower levels of corruption. High levels of corruption always go along with bad governance which then hinders economic growth by decreasing government revenue and increasing wasteful public spending. As a result, investment spending on education, a key factor for economic growth, reduces. This further lowers level of economic growth. In conclusion, the unique mechanism shared among all these four channels for a contribution to economic growth is through an increase in the level of economic efficiency.

In conclusion, the main basis for supporting fiscal decentralisation centres on the economic efficiency gains. Efficiency improvement will increase GDP. However, it is unclear whether this result is a one-off increase in GDP with no on-going increase in the growth rate, or

both a one-off increase and on-going increase in the growth rate. Static economic theory provides a compelling case for a one-off increase in GDP, the case for economic growth as well depends on dynamic issues and dynamic theory which is less well established because many different factors may influence to economic growth.

### **Governance Implications**

It is argued that fiscal decentralisation is a superior good which is suitable for the economy to adopt at a high level of economic growth and development. In this case, good governance is a main force to achieve effective and efficient fiscal decentralisation. However, this causal relationship is not secured. The reason for this view is that it should be possible for fiscal decentralisation to cause good governance since the decentralisation process brings with its many advantages. High level of fiscal decentralisation in society, by assigning powers and responsibilities to subnational governments to a great extent, could possibly enhance good governance through an increase in the level of local participation, economic efficiency, and government accountability. Experience has shown that, particularly for developing and transition economies, the process of decentralisation in the fiscal field has been continuously expanded. This reveals an agreement on the view that fiscal decentralisation does contribute to economic growth even though there may be the time lag. Further decentralisation of fiscal arrangements from high level to lower levels of governments is seen as a good opportunity for many economic and fiscal issues being sorted out, and then, in turn, contributing to economic growth.

### *Internet Sites*

Asian Resource Centre for Decentralisation.  
[www.decentralization.ws/srcbook\\_main.asp](http://www.decentralization.ws/srcbook_main.asp)  
 World Bank. *Decentralisation*.  
[www.decentralization.org/Active\\_Pages/index.asp](http://www.decentralization.org/Active_Pages/index.asp)

### **Selected References**

- Ahmad, Ehtisham and Craig, Jon (1997) "Intergovernmental Transfers" in Teresa Ter-Minassian (Editor), *Fiscal Federalism in Theory and Practice*. Washington DC: International Monetary Fund.
- Bird, Richard. (2000) "Intergovernmental Fiscal Relations: Universal Principles, Local Applications". *Working Paper No. 00-2*, International Studies Program, Andrew Young School of Policy Studies, Georgia State University.
- Bird, Richard. (1999) "Rethinking Subnational Taxes: A New Look at Tax Assignment". *IMF Working Paper*, Washington DC: International Monetary Fund.
- Boadway, Robin. (2003) "National Taxation, Fiscal Federalism and Global Taxation". The paper is prepared for The *WIDER Project on Innovation Sources of Development Finance*. Queen's University, Canada.
- Brennan, Geoffrey and Buchanan, James. (1980) *The Power to Tax—Analytical Foundations of a Fiscal Constitution*, Cambridge, UK: Cambridge University Press.
- Breton, Albert. (1996) *Competitive Governments: An Economic Theory of Politics and Public Finance*, Cambridge, UK: Cambridge University Press.
- Collins, David. (2001) "The 2000 Reform of Intergovernmental Fiscal Arrangements in Australia", Paper

- presented for the International Symposium in Quebec, Canada in September 2001.
- Dafflon, Bernard and Vaillancourt, Francois. (2005) *Problems of Equalisation in Federal Systems*. [www.federalism2005.be/home/attachme nt/i/621](http://www.federalism2005.be/home/attachme nt/i/621).
- Martinez-Vazquez, Jorge. (2001) "The Assignment of Expenditure Responsibilities", Document prepared for *Fiscal Policy Training Program 2001*. Washington DC: World Bank Institute.
- McLure, Charles. (2001) "The Tax Assignment Problem: Conceptual and Administrative Considerations in Achieving Subnational Fiscal Autonomy", Document prepared for *Fiscal Policy Training Program 2001*, Washington DC: World Bank Institute.
- McLure, Charles. (1998) "The Revenue Assignment Problem: Ends, Means, and Constraints", *Journal of Public Budgeting, Accounting and Financial Management*, Volume 9, Number 4, pp. 652-683.
- McLure, Charles and Martinez-Vazquez, Jorge. (2004). *The Assignment of Revenues and Expenditures in Intergovernmental Fiscal Relations*. Washington DC: World Bank.
- Musgrave, Richard. (1983) "Who Should Tax, Where, and What?" in *Tax Assignment in Federal Countries*, McLure, Charles (Editor), Australian National University, Canberra, Australia.
- Musgrave, Richard. (1959) *The Theory of Public Finance—A Study in Public Economy*. McGraw Hill Book Company, Inc.
- Oates, Wallace. (1972) *Fiscal Federalism*. Harcourt Brace Jovanovich, Inc.
- Organisation for Economic Cooperation and Development. (1999) *Taxing Powers of State and Local Governments*. Paris: OECD Tax policy Studies.
- Prud'homme, Remy. (1995) "The Dangers of Decentralization". *World Bank Research Observer*, Volume 10, pp. 201-220.
- Shah, Anwar. (2004) *Expenditure Assignment in Intergovernmental Fiscal Relations and Local Financial Management Program*. Washington DC: World Bank.
- Ter-Minassian, Teresa. (1996). *Borrowing by Subnational Governments: Issues and Selected International Experiences*. Washington DC: IMF paper on Policy Analysis and Assessment.
- Tiebout, Charles. (1956) "A Pure Theory of Local Expenditures", *Journal of Political Economy*, Volume 64, pp. 416-424.

Duc Hong Vo  
Department of Treasury  
Western Australian Government  
Perth, Australia

## Fiscal Policy

Leonardo V. Vera

### **Introduction**

Fiscal Policy generally refers to the use of taxation and government expenditure to regulate the aggregate level of economic activity and to moderate the economic fluctuations in a given economy. Indeed, the very term ‘fiscal policy’ came to mean the use of fiscal instruments as tools of macro-policy. But before the 1930s fiscal policy was not part of the lexicon of economists and the public finance inherited knowledge only dealt with selective market failures in the provision of public goods. The role of the State in ensuring the stability of the macroeconomy, in terms of employment, output and prices, begun to find recognition in the wake of the Great Depression experienced by the industrial economies in the 1930s. Since then economists’ view of fiscal policy usefulness has fluctuated widely.

Before the 1920s fiscal policy and specifically government spending policy was discussed under the name ‘sound finance’. The principle of sound finance maintained that the government budget should be balanced except in wartime. The classical liberal tradition in which this principle was held viewed government with suspicion, so any policy that would make it easier to increase government spending during peacetime was seen as undesirable.

But in the 1920s in Europe, and the 1930s in the United States, very well known economists, such as Arthur Pigou and John Maynard Keynes (in Cambridge, England) and Frank Knight, Paul Douglas, Henry Simons (at the University of Chicago), started questioning sound-

finance principles as the economies of the world fell into a major ongoing depression. Moreover, during the 1930s liberal business leaders share the feeling that running government budget deficits could help stimulate demand and they saw this type of action as a way to save capitalism with the least amount of government intervention (Colander & Matthews 2004).

Though an early generation of Chicago economists argued in favor of the use of the fiscal policy to absorb changes in the business cycle, the reasoning behind this view rested in the fragile financial structure that prevailed in the industrial economies at that time, in the impotency of monetary policy and in the presence of price and wage rigidities. The Chicago economists recommended injecting purchasing power through fiscal means that could restore the level of profitability (Perez 2003). Over time, as the belief that the economy operated close to full employment levels of output gained prominence among them and their followers, they abandoned the idea of reflationary or countercyclical fiscal policies. Keynes’ initial support of fiscal policy predated the *General Theory of Employment, Interest, and Money* (GT) and consequently his views in that respect were not based on it. Since the time of “*Does Unemployment Need a Drastic Reform?*” in 1924 and “*Can Lloyd George Do it?*” in 1929, Keynes had advocated public works, government expenditure undertaken in special circumstances as distinguished from regular expenditures, as a way to provide a palliative measure for the practical consequences of laissez faire. But it was later in *The Means to Prosperity* (in 1931) that he argued for public works programs using the logic of the Kahn’s multiplier. Keynes found no conflict between policies for increasing

employment and schemes for balancing the budget. On the contrary, according to Keynes striving to balance the government's budget during a slump would make things worse, not better. By increasing employment and national income the budget get balance. Indeed Pérez (2003) identifies this pamphlet as the first place where Keynes suggested the use of deficit spending as a stabilization tool.

Chick (1984) and Colander and Matthews (2004) correctly point out that Keynes' *GT* was not about fiscal policy at all. In the *GT* there is really no trace of deficit spending as a policy recommendation or as a policy tool. Certainly, Keynes argues, in one of his most controversial propositions, for a "comprehensive socialization of investment" but the discussion was far from complete. Keynes (1936) conceived private investment as driven by subjective evaluations of future profits and as inherently volatile, therefore; he concluded "that the duty of ordering the current volume of investment cannot safely be left in private hands" (p. 320). Thus, public, or publicly-guided investment, was required to offset the fluctuations of full employment and maintain full employment.

In the discussion of post-war employment policies and consistent with his emphasis on counter cyclical public investment, Keynes drew a sharp distinction between stabilizing investment and stimulating consumption by public action. He argued for a separation of the current or ordinary budget and the capital budget (Camara & Vernengo 2004). The first, according to Keynes, was related to government consumption and should be balance on average or even in surplus to finance the capital one. The capital budget

was related to public investment and should be used for counter cyclical purposes. Though anti-cyclical budgetary actions would be accomplished by changes in the capital budget, Keynes thought that it had to be balanced in the long-run (Guger & Walterskirchen 1988).

Though after WWII Keynesianism came to be considered as identical with fiscalism and fiscal policy became a form of demand management in a number of countries, we should say that in Keynes' thinking, fiscal policy did not occupy as central position as it did in the conventional Keynesian models and in policy-making in the 1950s and 1960s. Indeed Keynes' policy advice turned out to be much less straightforward than Lerner's concept of 'functional finance' which came to dominate fiscal policy in the early post-war period.

In proposing the rules of functional finance, Lerner's stated purpose was to show that the primary question facing governments is how to ensure that the impact of fiscal policy could be most beneficial upon the level of output and employment in the economy regardless of whether it increases or decreases government debt. Though the accumulation of government debt by current deficit-spending imply that future generations would be burdened with its payment, functional finance meant that the national debt is not a burden on posterity because posterity pays the debt when fiscal policy decisions affect levels of economic activity. This emphasis in output and employment regardless the level of government debt is consistent with the Chartalist view about how government fiscal policy operates. Accordingly, taxes do not, and cannot, finance government spending since the government cannot



collect more taxes than it has spent in one given year. If however the government were to limit its spending below the level that would provide citizens with the means of tax payment desired, there could remain an excess supply of goods and services to government and/or markets. From this viewpoint, resources are effectively unemployed because of insufficient government spending which is the source of means of tax payment.

Most industrialized economies soon absorbed Lerner's functional finance ideas and, for most of the post-war period, government fiscal and monetary policies were indeed assessed largely in terms of their effects on output and employment, regardless of whether these increased or decreased the public debt. This is precisely what is done in standard IS-LM exercises: the purpose of taxation or government spending is to shift the IS curve around and of increasing/decrease the supply of money and bonds is to move the LM curve so that output and employment will be at full employment—and not because it is necessary to 'raise revenue', 'raise funds', 'close the deficit', etc.

At the beginning of the 1960s, in the U.S. Kennedy's tax reductions to weaken 'fiscal drag' and later on the Vietnam war acted as expansive global forces. Furthermore, for some twenty years after the Second World War, Keynesian economic policies in countries of the capitalist West were successful in generating rapid growth with high employment. Cornwall and Cornwall (2005), for instance, take a close examination of the data for the post-World War II period (1960-1973) for the main 18 OECD countries and show that using a 3 per cent rate as the full employment rate of unemployment all but four of the eighteen

countries experienced full employment. They also note that during this period the low unemployment economies did not experience appreciably higher rates of inflation, indicating that the inflation rate was not merely a politically acceptable trade-off; and that these economies did not pay a higher inflation cost for their full employment.

This so-called 'golden age of capitalism' did not survive the economic traumas of the 1970s. When inflation in industrial countries went on rising after unemployment had stop falling, the composite model (the Keynesian/Phillips model) that appeared to be working satisfactorily in empirical terms until the second half of the 1960s proved to be inadequate. By the late 1960s and early 1970s the Keynesian remedies to ensure the stability of the macro-economy met with challenges from several quarters, primarily the monetarists. The resurgence of the quantity theory of money—under the new name of 'monetarism'—brought with both a renewed belief in the power of monetary policy and a resurgence of interest in the so-called 'crowding out' effect. Thus, monetarists clung to the view that fiscal policy was powerless.

There are several channels through which this crowding-out effect can occur. The most conventional form of crowding-out occurs when the deficit is financed by selling bonds. The price of the bonds is bid down (due to oversupply) which is equivalent to an increase in interest rates. The higher interest rate causes private investment to decline or to be crowded out as a result of higher deficit. An interesting development of the crowding-out hypothesis that emerged in the 1970s and has since been associated with New Classical Economics was the proposition of



the ‘Ricardian Equivalence’. Its origin dates back to the classical economist David Ricardo (1772-1823). In 1917 Ricardo wrote about how to finance a war with annual expenditures of \$20 millions and asked whether it makes a difference to finance the \$20 millions via current taxes or to issue government bonds with infinite maturity (so-called consols) and finance the annual interest payments of \$1 million in all future years by future taxes (at an assumed interest rate of 5%). His conclusion was that in the point of the economy, there is no real difference in either of the modes. He argued that in both situations taxpayers would have more money now; but they would realize that they would have to pay higher tax in future and therefore save the extra money in order to pay the future tax. Barro (1974) revived interest in the Ricardian equivalence proposition under the heading of *are government bonds net wealth?*

He developed more sophisticated variations on the same idea, particularly using the theory of rational expectation and a number of restrictive assumptions (people live forever, perfect capital markets, intergenerational altruism, lump-sum taxes and full employment). In essence, Barro (1974, 1989) argues that the effect of expansionary fiscal policy on future taxes leads consumers to change their saving. Recognizing that a tax cut today means higher taxes in the future, people will simply save the value of the tax cut they receive now in order to pay those future taxes. Therefore expansionary fiscal policy and an increase in debt cannot stimulate aggregate demand, and as a result, the increase in debt has no real effects.

The case for using discretionary fiscal policy to stabilize business cycles found

new challenges when analyzed in the context of developing countries. After the Second World War, governments in developing countries (in the initial years of independence as for instance in African countries or involved in some kind of planned development as in Latin America) had to spend a large amount of its resources in establishing the infrastructure in its broadest sense, viz., educational, financial, physical, technological and social. These countries had negative public savings and limited private investment. Despite large inflows of foreign aid, the increasingly large financing gap became the main concern of most government. The situation was further aggravated by the high exposition to shocks. Moreover, political incentives and external and fiscal constraints created a bias towards both high debt and high inflation. This gradually led to a weak fiscal structure and poor fiscal management raising vulnerabilities against the frequent internal and external shocks.

The policy approach that would eventually become the mainstream solution to these multifaceted problems in developing countries in the early 1980s was explicitly hostile to State intervention in the economy. The particular policies that were suggested—and then imposed in the context of international agreements—that eventually became known as the Washington Consensus (Williamson 1990) included as a first item fiscal discipline. In some sense it was a return to the principles of sound finance. The emphasis on fiscal discipline was associated with the concern that high fiscal deficits were behind macroeconomic instability, on one hand, generating inflation, and, on the other, generating balance-of-payments problems. Unfortunately, under the influence and

pressure of the IMF/World Bank adjustment programs, monetary and fiscal policy management became even more procyclical, deepening the cycle particularly during recessions. A procyclical bias in fiscal policy occurs when, during recession, government spending decreases and taxes goes up. There are a limited number of empirical studies that show that fiscal policy has been historically strongly procyclical in developing countries (see, for instance Kaminsky et al 2004; Braun 2001; Gavin & Perotti 1997; Akitoby et al 2004).

As emphasized by Perry (2003), this procyclical character of fiscal policy may end up generating unsustainable fiscal results over the cycle—as proved to be the case in several Latin American countries, most recently in Argentina and Ecuador. This may explain also why most of the discussion on fiscal policy in developing countries turned later to long term sustainability issues, largely ignoring the short-run effects on economic fluctuations and employment.

Arguments against State intervention in the economy of course had strong contemporary overtones, especially in the context of the American and British economy of the early 1980s. The inflation rate of the late 1970s was enough to convince President Reagan that the most acceptable measure against inflation seemed to be a curb of government expenditures. The subsequent 1981 tax cut (a landmark in the history of fiscal policy) was also justified not by the Keynesian aggregate demand considerations—which were denigrated—but a new doctrine called supply-side economics or the economics of incentives. From a fiscal view, tax cuts were supposed to produce higher tax collections. Although the whole approach

turned out to be about reducing the presence of the State in the economy as well as the fiscal deficit, the expectations were in blatant contradiction with later events. A buildup of defense expenditure, continued growth in spending on entitlement programs, such as Social Security and Medicare, and an effective lower tax collection combined to produce a huge deficit. Moreover, the period 1981-88 was one of recovery from the recession, bringing unemployment back down to 6%. This is the reason why American Nobel prize James Tobin understood Reaganomics as “the biggest and most successful demand-side fiscal gambit in peacetime U.S. history” (Tobin 2001:3). The deficit spiked early to reach 6.2 point of GDP in 1983 and 5 points in 1984, then it declined gradually in the second term of Reagan's presidency. In Britain, the advent of the Thatcher years heralded a major departure from State intervention which was accompanied by the downgrading of fiscal policy and the return to balanced budgets.

The Reagan legacy of huge deficits fostered a dramatic repositioning of fiscal policy in the U.S.. Blinder (2004) finds a new devotion to fiscal prudence, in 1985, when Congress passed the Gramm-Rudman-Hollings Act which required adherence to annual targets for the federal budget deficit. A similar trend in Europe established a number of aims for macroeconomic policy specified in the Treaties around the European Monetary Union (EMU). The EMU was based on an original arrangement of public finance relations between member countries where fiscal policy remained decentralized, but was subject to rules which were meant to combine discipline and flexibility. The Stability and Growth Pact (SGP), which

complemented and tightened the fiscal provisions, laid down in the Maastricht Treaty (the backbone of fiscal discipline in EMU). The SGP is widely viewed today as the most stringent “commitment technology” ever adopted by sovereign governments on a voluntary basis in the attempt to establish and maintain sound public finances. The fiscal rules of EMU were based on a simple predicament: government should reduce budget deficits to close to balance and then let automatic stabilizers play freely. Discretionary counter-cyclical budgetary policy was not really considered integral to this system. This seems to be surprising to some extent since monetary policy within a monetary union can no longer play this role. As argued by Buti and van den Noord (2004) EMU is commonly seen as a regime of monetary leadership where the role for fiscal policy is to support the central bank in its task to keep inflation in check.

The view among academic circles and policymakers regarding the usefulness of fiscal policy may have swung just a bit too far in the 1980s and 1990s. As Blinder (2004:1) points out “virtually every contemporary discussion of stabilization policy by economists—whether it is abstract or concrete, theoretical or practical—is about monetary policy, not fiscal policy”. Moreover, the high world economic instability that prevailed in the last two and a half decades led most national governments in emerging and developing economies to implement fiscal policies targeting higher deficits. Thus, the budget deficit has held a role of great, even mythic, importance in public discussions. However, although the ‘heritage’ of the debate in the 1980s and 1990s casts a strong skepticism over the use of discretionary fiscal action to fine tune the

economy, the potential usefulness of fiscal stabilization is being re-considered. The debate on fiscal policy has revived, for instance, in Europe where it now centers on how to facilitate the workings of automatic stabilizers as a counter cyclical instrument, or in Japan where the prolonged slump have seen a revival of the debate about its effects in stimulating economic activity.

### **Relevance of Counter Cyclical Fiscal Policy**

In the short run, the possible role that fiscal policy could play in stabilizing output may occur through the operation of automatic stabilizers and/or discretionary fiscal policy. Automatic stabilizers are budget components that respond automatically to the business cycle without any explicit government action, while discretionary policy consists of active policy measures meant to stimulate the economy during bad times. The appropriateness and feasibility of either, as we will see, may vary according to the individual country circumstances.

The popular view on automatic stabilizers relies on the assumption that fluctuations in GDP or income are partially smoothed by changes in taxes and transfers over the business cycle so that disposable income is less volatile than income. As the economy slides into a recession incomes are falling but collected income taxes falls, unemployment is rising but transfers and payments of unemployment benefits also rises. In this setting automatic stabilizers have obvious appeal as a counter cyclical policy instruments since they are supposed to be not subject to time inconsistency problems. These arrive when government policymakers are best off to promise that there will be no inflation tomorrow, but

once agents and firms in the economy have fixed nominal contracts, the government decides to expand public expenditure. In opposition to discretionary actions, automatic stabilizers are not affected by implementation lags. Precisely because they are not discretionary, they are also less likely to affect market expectations adversely. It is also argued that automatic stabilizers are more effective in stabilizing output fluctuations because they are more predictable and, unlike discretionary measures, they do not require political forecasting.

But there are a number of factors that may account for the weak automatic stabilizers. For example, automatic stabilizers may be constrained by the combination of low tax elasticity and relatively low share of taxes in GDP that tends to reduce the responsiveness of revenues to demand shocks. Their impacts depend on how strongly consumption demand responds to changes in disposable income and this in turn may depend on whether the shock is seen to be temporary or permanent. They are relatively ineffective when the source of the shock to the economy is from the supply rather than the demand side and since they are backward-looking by nature, they are less useful in preventing a demand shock. Furthermore, it is likely that automatic stabilizers are less important in developing countries. In these economies the revenue/GDP and expenditure/GDP ratios are far smaller than in advanced countries. Since the tax base in these economies is small, the share of income-elastic taxes is smaller than in industrial countries. On the expenditure side, few developing countries have significant social security, and unemployment benefits are not an important expenditure category that moves

with the cycle. At the same time, improving automatic stabilizers poses an important challenge: it implies introducing additional welfare and unemployment programs, which countries may be unable to afford without raising their fiscal deficits.

The above discussion highlights the fact that there are circumstances and contexts where discretionary counter cyclical fiscal policy could be a far better candidate for the stabilization job. This is clearly the case in Southeast Asian economies, where the recession that followed the Asia crisis was attacked with expansionary fiscal policy (Jansen 2002). But many analysts have argued that discretionary fiscal policy is especially difficult to use for stabilization because of the ‘inside lag’—the gap between the time when the need for fiscal policy arises and when it is implemented by the executive and Congress. Arestis and Sawyer (2004) clearly remark that fiscal policy is much more subject to democratic decision-making than monetary policy. Thus, what seems to be its strength in terms of social and political consensus, can also be its weakness in terms of its economic goals. But lags are not immutable. The sources of many, if not most, of them lie in policymaking institutions that can be changed (Blinder 2004). On such idea is to reduce the inside fiscal lags by the adoption of a ‘fiscal policy rule’ (Taylor 2000; and Budnevich 2002) so long as it emphasizes full employment (Arestis and Sawyer, 2004) and where authorities should respond to output fluctuations through a fiscal policy reaction function.

Another question that arises with respect to discretionary fiscal policy is that it is sometimes very difficult to implement in financially open economies where a

fixed exchange rate system prevails. This is precisely the case of many developing economies. Under borrowing constraints, an adverse external shock sometimes requires a reduction of absorption, making fiscal policy contractionary. Moreover, orthodox economists and multilateral agencies such as the IMF argue that, against an adverse external shock, fiscal contraction gives international investors confidence avoiding capital outflows. To this there are others that argue that investors will lose confidence when they see that fiscal and monetary contraction leading the economy into a deeper recession.

Another additional issue arises from the alleged possibility of crowding-out. This already described view on the effectiveness of fiscal policy is quite controversial. In particular what seems to be restrictive is to neglect a possibly complementary relationship between public and private investment. Keynes sharp distinction between stabilizing investment and stimulating consumption by counter cyclical public action is quite relevant here. Public and private investment may be linked by a complementarity relationship if public capital provides positive externalities on the private sector. It was Aschauer (1989), who in a seminal empirical contribution, showed that for the United States, there was a strong positive relationship between productivity and the ratio of the public investment to the private capital stock. After that, several authors have elaborated on the work of Aschauer and have look at the issue for other developed and developing countries.

In this so-called crowding-in hypothesis many channels may be involved. Belloc and Vertova (2006) summarize three of them. With the first, the availability of

economic and social infrastructures may create favorable conditions for private decisions to invest, by offering essential services to the production system both in the short and long run (transportation, communication, education, and so on). Secondly, higher public capital may lead, on the one side, to increments in total factor productivity and, on the other, to reductions in production costs (through availability of streets, highways, electrical and gas facilities, mass transit, and so on). And finally, public investment, by increasing total demand, may give rise to profit and sales expectations, so to spur private decisions to invest more.

Thus even if private investment is found directly more productive than public investment, any conclusion on adjustment strategies should be qualified with the consideration of the relationship between public and private investment. Indeed, if the crowding-in hypothesis holds, a fiscal adjustment which reduces public investment implies a contraction in the fixed capital formation and a slowdown in economic performance.

### **Fiscal Policy and Inflation**

There is a further issue around the idea that counter cyclical fiscal policy may entail a 'deficit bias'. For a long time, economists and policymakers have worried about the relationship between government budget deficits and inflation. The orthodox view has been that the main culprit behind the inflationary process is the creation of high-powered money (and thus seignorage) to help finance fiscal imbalances. Through the so-called 'seignorage', the government appropriates an amount of real resources by means of base money creation. Under certain circumstances if a larger budget deficit translates into a larger money stock

of money, then the money issuing authority will be in practice attempting to force agents to hold more real balances than they otherwise would. Inflationary expectations, and inflation, should then be adjusted upwards and real balances would fall further. This inflationary finance story, that occupies a distinguished place in descriptions and explanations of inflations and hyperinflations, was heavily influenced by the works of Philip Cagan (1956).

There are several reasons why the relationship between deficit financing and inflation is not straightforward or fails. One is that countries make different choices on printing money to finance the deficit, partly because they differ in the extent to which other means of finance are available, partly because the tax base (the stock of outside money) is often rather low. Some high-deficit countries finance the deficit with cheap domestic finance, some have access to extensive concessional external finance and some effectively eliminates printing money as a means of financing. To illustrate this we can use the following identity depicting the intersection of the government budget constraint and the balance sheet of the Central Bank, namely

$$(1) \quad (G - T) = (M_t - M_{t-1}) + (B_t - B_{t-1}) - e(R_t^* - R_{t-1}^*)$$

The fiscal deficit on the left-hand side depends on fiscal revenue,  $T$ , and global government expenditure,  $G$ . Thus a positive deficit will imply a change in the money stock,  $M$ , and change in the stock of public debt held by the public,  $B$ , or a change in the international reserves position,  $R^*$ . Each financing mechanism would entail different macroeconomic repercussions; money printing would be linked to inflation, use of reserves with exchange rate movements and possible

balance of payments crises, foreign borrowing with external debt crises, and internal borrowing with higher interest burden and potentially explosive debt dynamics.

A second reason that may invalidate the inflationary finance story has to do with the fact that a surprising number of episodes of high seignorage are due to increases in real money balances instead of accelerating inflation. Therefore, seignorage may increase even when inflation is nil, due to increases in the demand for money, for instance.

In an open economy, inflationary finance would exist insofar as a flexible exchange rate system is assumed. But the inflationary finance story also requires zero output growth, a constant velocity of money and fully flexible prices in the economy. Thus a deficit financed by money creation creates a situation in which agents find themselves holding excess money balances that they spend on foreign goods and/or external assets. As domestic residents sell their local currencies in exchange for foreign currencies, a nominal depreciation occurs. If the purchasing power parity holds then domestic prices will increase.

One important difficulty with the canonical inflationary finance approach in open economies arises from the fact that references to adverse shocks in the balance of payment are absolutely absent. In the orthodox approach the fiscal deficit would be the ultimate cause of the money creation and inflation. The prior question about what produces the fiscal deficit is left unanswered. But the fact of the matter is that reliance on fiscal austerity alone does not address the core of the problem. Heterodox economists of different persuasions have emphasized over the

years adverse external shocks such as war reparation payments (in the early 20s) or foreign debt services (in the 80s) as basic causes of the exchange rate devaluations and of inflation. Indeed, in the presence of distributive conflict, the fall of real wages following a real devaluation would be resisted through increases in nominal wages, and accommodated with greater inflation. In this context, money is passive in the sense that money supply is adjusted to the evolution of the exchange rate and prices in sustaining higher nominal income levels. Solimano (1989) and Camara and Vernengo (2001), for instance, provide analysis of these alternative or challenging views to inflation that surged within the context of the European hyperinflation of the 1920s and 1940s, and in Latin America after the several high inflation episodes of the 1980s.

### **Fiscal Policy and the External Sector**

Since 2002, the U.S. has seen the simultaneous emergence of current account and fiscal deficits which apparently reflects increasing U.S. borrowing from abroad. These deficits have been called twin deficits and the linkage is an automatic result of a national account identity in which injections equals leakages in terms of the circular flow of income.

$$(2) \quad (X - IM) = (S - I) + (T - G)$$

where  $S$ ,  $I$ ,  $X$  and  $IM$  stand for private saving, private investment, exports and imports plus net factor services to abroad respectively. The identity became commonplace during the 1980s and 1990s in the U.S. because it was supposed to entail a relationship between budget and trade deficits. The relationship has also been widely discussed in the context of

developing countries. The hypothesis states that a budget deficit caused, for instance, by an expansionary fiscal policy ( $G > T$ ) will lead to a current account deficit ( $IM > X$ ).

Standard economic theory would not find the situation surprising. A first theoretical explanation of the relation between fiscal policy and current account deficits can be found in the Mundell-Fleming framework (MF). The MF approach indicates that an increase in government expenditure or a decrease in taxation induces an upward pressure on interest rates that, in turn, will trigger capital inflows and an appreciation of exchange rates, ultimately leading to an increase in the current account deficit. A second theoretical explanation of the linkage between the twin deficits is the Keynesian absorption theory, which suggests that an increase in the budget deficit would induce domestic absorption, that is an increase of total expenditures of domestic residents, and hence, import expansion, causing an increase or a worsening of the current account deficit.

The experience, both in developed and developing countries, shows that the relationship between fiscal deficits and external balance takes a wide variety of forms that go beyond the simple twin deficit causality. It may happen, for instance, that the foreign account is balanced while the budget deficit has to match the borrowing of the private sector. It is perfectly possible also that a fiscal deficit may be caused in the first instance by an autonomous fall in exports, an autonomous increase in imports, or an autonomous rise in the international interest rate, quite independent of government decisions to spend. In commodity exporters and highly indebted



countries, for instance, what happens very often is that they face inherent instability from fluctuating export prices and foreign interest rates that deteriorate the current account and translates into higher budget deficits. If causality is reversed, then the policy implications are substantially different. Also, sometimes domestic investment could be the driving force, from the domestic side of the economy, leading to a deteriorating trade balance (Blecker 1999).

Moreover, budget and current account deficits sometimes follow quite divergent paths. One possible explanation for this divergence is related to the impact of output fluctuations on budget and current account deficits. For instance, if the economy enjoys a surge in productivity that prompts an expansion in economic activity, then to reap the opportunities of higher productivity, private investment increases. As investment expenditure typically reacts more strongly to the business cycle than private saving does, and the current account balance deteriorates. At the same time, the output expansion generates both an increase in tax receipts and a decline in government expenditure, due, for example, to a decline in unemployment benefits. Therefore, the budget balance improves.

Another explanation of the relation between budget and current account deficits is based on the presumption that the twin deficits are not related. Such a view is supported by the Ricardian Equivalence Hypothesis. It states that, for a given expenditure path, the substitution of debt for taxes has no effect on aggregate demand nor on interest rates. As a result, it implies that a tax increase would reduce the budget deficit but would not alter the external deficit, since altering the means

that the government uses to finance its expenditures does not affect private spending nor national saving. Empirical work on the Ricardian equivalence proposition in industrial countries produces evidence that is mixed (Fischer & Easterly 1990) while for developing countries is rather weak.

### **Sustainable Fiscal Stance**

Debt sustainability has become one of the most used and abused concepts in recent discussions regarding international financial issues. Public Debt sustainability is an important issue, in particular to the countries belonging to the EMU, where the need to ensure fiscal sustainability was often invoked as a rationale for the fiscal rules set out in the Maastricht treaty and in the SGP. Debt sustainability is also important in a number of developing countries, where the fiscal budget has been caught up in an explosive spiral of increasing indebtedness—a ‘snowball effect’ in which the debt generally absorbs a growing proportion of fiscal revenue.

Sustainability of the fiscal stance is a long-term problem and concerns the way in which budgetary viability may get eroded over time due to financing of government expenditure through borrowing. The most known concept of fiscal sustainability relates to the government’s ability to indefinitely maintain the same set of policies (regarding taxes and expenditure for instance) while remaining solvent. This means that the focus of fiscal sustainability analysis is frequently not on default itself—which governments frequently avoid—rather it is on the consequences of the policy changes needed to avoid eventual default. The notion of fiscal sustainability is not new. In the beginning of the 20s, for instance, when writing about the public



debt problem faced by France, Keynes (1923) alerted to the need for the French government to conduct a sustainable fiscal policy in order to satisfy its budget constraint. Keynes stated that the absence of sustainability would be evident when "the State's contractual liabilities ... have reached an excessive proportion of the national income" (p.54). In Keynes's words, there is a problem of sustainability when "it has become clear that the claims of the bond-holders are more than the tax payers can support" (p.55). According to Keynes, at that stage the government "must come in due course to some compromise between increasing taxation, and diminishing expenditure" (p.59). But Keynes was never in favor of repayment at whatever cost as he left it very clearly when denouncing the absurdity of the reparation question after the Versailles peace treaty (Keynes 1920).

In the early 1940s, Domar (1944) developed the well established, although presently a bit underrated, strand of the literature that identifies sustainability with the dynamic stability of the public debt/GDP ratio around a constant steady state. This definition has a lot intuitive appeal. Accordingly, the dynamic stability of the debt-GDP ratio is ensured as soon as the rate of output growth in the economy is greater than the real interest rate on public debt. In a further contribution, Blanchard (1990) devised a simple framework for the analysis of sustainability in which a budget debt will be sustainable at any point of time when the value of current debt is lower than the net present value of future primary balances, that is, the fiscal balance once the interest payments on debt are deducted. But this definition, simple as it is, has faced the problem of not being operative, since it is quite difficult to

derive the series of future fiscal balances or to impose a particular rate of discount on the future.

A basic problem with the conventional sustainability framework is the critical assumption about the behavior of key macroeconomic variables. A high and explosive debt/GDP ratio is supposed to be controlled by reducing government expenditure. However, as remarked by Jha (1994) this may not be as simple as it looks at first glance. Reducing government expenditure may lower real national income and then tax revenues and exacerbate the debt situation. Thus, using sustainability targets to correct the size of the primary balance (to avoid default) may not be such a good idea since government spending cuts may be self-defeating. Moreover, it is sometimes argued that a deficit that results from high public investment will be sustainable since spending on public infrastructure often promotes growth (though inefficient public investment has been extensively documented in developing economies). Indeed, as shown by Rakshit (2000), when the government borrows in order to meet capital expenditure, it accumulates assets. As a consequence, sustainability (in the Domar sense) now requires the growth rate to be higher than the interest rate less the return on assets being accumulated by the government—a much less onerous condition than the canonical one.

## **Conclusion**

Although the heritage of the debate in the 1980s and 1990s casts a strong skepticism over the use of discretionary fiscal action to fine tune the economy and stabilize business cycles, deeper inspection reveals that the relationship between fiscal policy and the rest of the economy has remained

one of great contested areas in macroeconomics. Indeed, economists' view of fiscal policy usefulness has fluctuated widely since 1920s.

There are good reasons to think that the potential usefulness of fiscal stabilization needs to be re-considered. Concerns about investment crowding out, inflation, scarce saving, and international trade problems do not pose insurmountable barriers to the effectiveness of fiscal policy. But since each economy is unique, none of the links between fiscal policy and the rest of the economy is automatic, for there are institutional and structural configurations that imply specific causality mechanisms as well as choices in policy instruments and sources of financing that may have different macroeconomic effects.

Further analysis is necessary to be able to provide guidance regarding the form to integrate both macroeconomic stability and solvency considerations. The current method of integrating them, instituting restrictions on deficits, works pro-cyclically, and often worsens the fluctuations. The pro-cyclical character of fiscal policy is an increasing burning topic and the economic solutions in vogue today are biased in favor of the use of rules. But more research is needed here since any rule may entail a dilemma between flexibility and credibility and a too rigid one, in the pursuit of credibility, may lead to high cost in forgone flexibility.

Increasing interdependent and integrated markets presents a country's economy with a number of dilemmas as well. Changes in the external environment may improve or worsen the domestic fiscal situation. However, in developed countries the changes in the fiscal variables can be attributed to policy responses (an adverse oil price shock, for instance, reduces real

incomes but may face a conscious discretionary government reaction) while in developing economies, the observed change in the fiscal variables may not be attributed mainly to policy changes but to the link that exists between the budget and the foreign sector. Several questions need to be addressed in this respect: what is the structure or institutional arrangement which a country's economy needs in order to be able to pursue an independent fiscal policy? Is this in fact possible in an increasing global economy?

Though fiscal policy remains a powerful instrument for regulating aggregate demand when the economy's resources are underutilized, research on the combination of public spending and revenues that is most helpful to achieve higher output and poverty reduction is required. Moreover, the reader is not necessarily forced to accept the view that fiscal policy must face an immutable barrier by full employment. Recently, new endogenous growth literature have proposed a number of channels through which fiscal policy could induce increases in potential output (affecting factor accumulation or influencing technical progress). This is an area where the identification of the required fiscal policy instruments and prescriptions is needed.

### **Selected References**

- Akitoby, B.; B. Clements, S. Gupta and G. Inchauste. (2004) *The Cyclical and Long-Term Behavior of Government Expenditures in Developing Countries*, IMF Working Paper Number WP/04/202, Washington, D.C., International Monetary Fund.
- Arestis, P. and M. Sawyer. (2004) *Fiscal Policy: A Potent Instrument*, *The New*

- School Economic Review*, Volume 1, Number 1, pp. 21-32.
- Aschauer, D. (1989) Does Public Capital Crowds Out Private Capital?, *Journal of Monetary Economics*, Volume 23, pp 177-200.
- Barro, R. (1974) Are Government Bonds Net Wealth?, *Journal of Political Economy*, Volume 82, Number 6, pp. 1095-1117.
- Barro, R. (1989) The Ricardian Approach to Budget Deficits, *Journal of Economic Perspectives*, Volume 3 Number 2, pp. 37-54.
- Belloc, M. and P. Vertova. (2006) How does Public Investment Affect Economic Growth in HIPC? An Empirical Assessment, *International Review of Applied Economics*, Volume 20, Number 2, pp. 151-170.
- Blanchard, O. (1990) *Suggestions for a New set of Fiscal Indicators*. Working Paper 79, OECD, Economic Department.
- Blecker, R. (1999) *The Causes of the U.S. Trade Deficit, Statement before the Trade Deficit*. Review Commission, Washington D.C.
- Blinder, A. (2004) *The Case Against the Case Against Discretionary Fiscal Policy*. CEPS Working Papers Number 100, Princeton University.
- Braun, M. (2001) *Why is Fiscal Policy Procyclical in Developing Countries?* Mimeo, Harvard University.
- Budnevich, C. (2002) Countercyclical Fiscal Policy: Review of the literature, empirical evidence and some policy proposals, *Discussion Paper* Number 2002/41, World Institute of Development Economics Research, Helsinki.
- Buti, M. and P. Noord. (2004) Fiscal Policy in EMU: Rules, Discretion and Political Incentives, European Commission, Directorate-general for Economic and Financial Affairs, *Economic Papers*, Number 206.
- Cagan, P. (1956) "The Monetary Dynamics of Hyperinflation", in M. Friedman (Editor), *Studies in the Quantity Theory of Money*. Chicago: Chicago University Press, pp. 25-117.
- Camara, A. and M. Vernengo. (2004) Fiscal Policy and the Washington Consensus: A Post Keynesian Perspective, *Journal of Post Keynesian Economics*, Volume 27, Number 2, pp. 333-343.
- Chick, V. (1984) *Macroeconomics After Keynes*, The MIT Press, Cambridge Massachusetts.
- Colander, D. and P.H. Matthews. (2004) Integrating Sound Finance with Functional Finance. *Middlebury College Economics Discussion Paper* Number 04-13, Department of Economics, Middlebury College.
- Cornwall, W. and J. Cornwall. (2005) A Future for Keynesian Macroeconomics. *CEPP Working Paper* Number 01/05. Cambridge, UK: University of Cambridge.
- Domar, E. (1944) The 'Burden of Debt' and the National Income, *American Economic Review*, Volume 34, pp. 798-827.
- Fischer, S. and W. Easterly. (1990) The Economics of the Government Budget Constraint, *World Bank Research Observer*, Volume 5, Number 2, pp. 127-42.
- Gavin, M. and R. Perotti. (1997) Fiscal Policy in Latin America, *NBER Macroeconomics Annual 1997*, Volume 12, pp. 11-71.
- Guger, A. and E. Walterskirchen. (1988) "Fiscal and Monetary Policy: The Keynes-Kalecki Tradition" in J. Kregel, E. Matzner and A. Roncaglia (Editors),

- Barriers to Full Employment*. New York: St. Martin Press.
- Jansen, K. (2002) The Scope for Fiscal Policy: With examples from Thailand, *Working Paper* Number 369, Institute of Social Studies, The Hague.
- Jha, R. (1994) *Macroeconomics for Developing Countries*. London: Routledge.
- Kaminsky, G., C. Reinhart, and C. Végh. (2004) *When it Rains, it Pours: Procyclical Capital Flows and Macroeconomic Policies*. NBER Working Paper Number 10780.
- Keynes, J.M. (1920) *The Economic Consequences of the Peace*. New York: Harcourt, Grace and Howe.
- Keynes, J.M. (1923) *A Tract on Monetary Reform*, in *The Collected Writings of John Maynard Keynes*, Volume IV, London: Macmillan, 1971.
- Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*. London: Macmillan.
- Perez Caldentey, E. (2000) "Chicago, Keynes and Fiscal Policy", *Investigación Económica*, Volume 62, Number 246.
- Rakshit, M. (2000) *On Correcting Fiscal Imbalances in the Indian Economy: Some Perspectives*. ICRA Bulletin Money and Finance, July-September.
- Solimano, A. (1989) *Inflation and the Costs of Stabilization: Country Experiences, Conceptual Issues and Policy Lessons*. Working Papers Series 226. Country Economics Department, Washington DC: World Bank.
- Taylor, J. (2000) Reassessing Discretionary Fiscal Policy, *Journal of Economic Perspectives*, Volume 14, Number 3, pp. 21-36.
- Tobin, J. (2001) Fiscal Policy: Its Macroeconomic Perspective. Cowles Foundation Discussion Paper Number 1301. New Haven, Yale University.
- Williamson, J. (1990) "What Washington Means by Policy Reform", in J. Williamson (editor), *Latin American Adjustment: How Much has Happened?* Washington D.C., Institute for International Economics.

Leonardo V. Vera  
School of Economics  
Universidad Central de Venezuela  
Caracas, Venezuela  
leonardo.vera@cantv.net

## Fiscal Sociology

*Michael McLure*

### **Introduction**

Analysis of social influences on public finances is generally referred to as fiscal sociology. Modern fiscal sociology is largely a legacy of pioneering investigations, undertaken towards the end of World War I, by Rudolph Goldscheid and Joseph Schumpeter on the question of whether fiscal crises necessarily emerge in states that rely on general taxes to fund the provision of public services (i.e. ‘tax states’). While the issue of fiscal crisis has remained a significant concern in modern fiscal sociology, the scope of the field has expanded to include constitutional, legal and institutional influences on fiscal policy, including the influence of bureaucratic agendas of fiscal decisions.

### **Emergence of Fiscal Sociology**

Fiscal sociology emerged as a field of study in Europe, with two distinct traditions developing, one in Austria and the other in Italy. Within the Austrian tradition, seminal contributions were made by Rudolph Goldscheid, such as in his 1917 *Staatssozialismus oder Statkapitalismus* and 1925 “*Staat, öffentlicher Haushalt und Gesellschaft, Wesen und Aufgaben der Finanzwissenschaften vom Standpunkte der Soziologie*” (1925), and by Joseph Schumpeter, especially in his influential “*Die Krise des Steuerstaats*” (1918). Within the Italian tradition, formative contributions were made by Gino Borgatta, in works such as “*I problemi fondamentali della scienza finanziaria*” (1913) and “*Lo studio scientifico dei fenomeni finanziari*”

(1920) and Guido Sensini in his “*Cenni di finanza teorica*” (1929).

The Austrian approach recognised that the ‘sociology’ of public finances, developed from an historical investigation of fiscal information, provides considerable insight into both the fiscal activities of government and the prevailing social structures and arrangements. Goldscheid and subsequently Schumpeter placed the notion of the ‘tax state’ as the subject of their analysis, with financial history and the consequent development of public debt and fiscal crisis central to their critical assessments. Notwithstanding their agreement on applying sociology to the study of fiscal matters, the differences between these two approaches were significant, with Goldscheid and Schumpeter ultimately reaching quite different conclusions on the nature and sustainability of the tax state.

Goldscheid was influenced by Karl Marx, and relied on class analysis to analyse the exploitative aspects of fiscal outcomes. From fiscal history, Goldscheid found that States progressively disposed of public property in response to increased enfranchisement, and provided the opportunity for public creditors—whose interests he equated with capitalists and the ruling class—to exploit the state as it came to progressively rely more and more on debt. In this circumstance, tax pressures to service debt was presented as a manifestation of social struggle, with fiscal exploitation of workers by the tax state (for the benefit of public creditors) being interdependent with, and complementary to, capitalist exploitation of workers throughout the private economy. A tendency to fiscal crisis became a consequence of the relationship between the fiscal system and capitalism.

In contrast, Schumpeter analysed the nature of the tax state in terms of an evolution in the complexity of the relationship between individual and collective needs. He reserved use of the term ‘the state’ to circumstances where a ‘public purpose’, as opposed to an individual purpose, was pursued on behalf of the collective. While “Goldscheid saw the emergence of the tax state as leading to state impoverishment, Schumpeter heralded it as the very creation of the State” (Musgrave 1992:33). Furthermore, while Goldscheid utilized class analysis, Schumpeter tended to emphasize groups and the role of interests in shaping the activities of the State: “It is always important to recognize who or whose interest it is that sets the machine of state in motion and speaks through it.” (Schumpeter 1918:19). Consequently, Schumpeter’s notion of ‘common’ need was not naive or simplistic, but was considered as something that emerged within the general context of the relationships of social power that exist within and between groups.

Goldscheid’s solution to the twin issues of fiscal crisis and fiscal exploitation was for the working class to conquer the bourgeois classes by creating a wealthy State (a ‘state capitalism’), with the State acquiring property to use for economic purposes and for the associated ‘surplus value’ to be expended in the interests of society. For Schumpeter, however, the case for fiscal crisis had not been demonstrated. His interest was not in periodic or avoidable crises, but the fundamental question of whether the tax state must descend into fiscal crisis as a result of funding the state provision of services in response to a mixture of common need and private purpose. By complementing his

sociological explanation of the growing demand for public expenditures with an economic explanation of the relative stagnation in tax receipts (Seidl 1984:91), Schumpeter was able to conclude “there is no ‘crisis in the tax state’” (Schumpeter 1918:37). That is, while tax states can collapse, a collapse is not a logical necessity of the tax state, at least not in the particular historical moment under consideration.

The Italian approach to fiscal sociology has at least one common factor—its main proponents were ex-students or associates of Vilfredo Pareto. A change in fiscal settings modifies the existing combination of revenue raising, expenditure and debt which, according to Pareto, cannot be considered in isolation. Instead, the variation in economic and social equilibrium resulting from each aspect of a new fiscal setting must first be established before arriving at any judgement on the welfare implications of the new fiscal settings. In a letter written in September 1917 to Benvenuto Griziotti, Pareto complained that “the ‘science’ of finance knows little of economic equilibrium and nothing of social equilibrium” (Pareto 1975:984), but confessed that he was unsure of how fiscal studies should be progressed, contenting himself with the observation that “... at least I have the merit, like Socrates had, of knowing my own ignorance” (Pareto 1975:984).

Gino Borgatta and Guido Sensini were the most prominent of Pareto’s followers to have contributed to the development of Italian fiscal sociology. The importance of these authors is that they developed an analytical framework to accommodate an economics and sociology of public finance that focused on social equilibrium. The starting point of their sociology was that

public finances provide a conduit for redistributing economic goods among individuals and groups, altering economic activity in the process. In recognition of this, Italian fiscal sociology examined the conditions that determine the form of a society as well as the interacting factors that have the potential to alter on the degree of stability in society's prevailing social equilibrium. Such investigations evolved from a Paretian general sociology into a genuine fiscal sociology, because Borgatta and Sensini recognized that public finances play an integral role in realizing the general regularities associated with a particular social equilibrium. In this regard, while the main attributes of social utility within a given society may be identified through an applied sociology, the fiscal dimension emerges as significant when considering the type of economy within a particular society, the economic and social characteristics of the governing class, the stability of the relationship between the governing class and the governed and the stability of the relationship between the governing class and the economy.

When reflecting on developments in Italian fiscal studies to the middle of the Twentieth Century, Benvenuto Griziotti (1950) concluded that the renewal of fiscal studies sought by Pareto was largely complete, with fiscal research in Italy developing along three related but distinct lines: (i) the science of public finance; (ii) the macroeconomics of public finance; and (iii) the sociology of public finance. Somewhat surprisingly, the Italian approach has had little influence on modern fiscal sociology. No doubt language barriers have contributed to this unfortunate neglect, although it is possible that the concurrent developments of fascism and fiscal sociology in Italy may

have also contributed the lack of enthusiasm for the subject.

The international dissemination of essentially European studies in fiscal sociology dates back to the 1940s and the publication of essays by Fritz Karl Mann (1943, 1947), an exile from Nazi Germany, in prominent English language journals. However, it was not until the translation and publication of Schumpeter's "Crisis of the Tax State" (Schumpeter 1954) and Goldscheid's "A Sociological Approach to Problems of Public Finance" (Goldscheid 1967) that the subject took a firm hold within America. Perhaps not surprisingly, the classic issue of fiscal crisis became the first theme to be treated within American fiscal sociology.

### **American Fiscal Sociology of Crisis**

The American reprise in the sociology of fiscal crisis is primarily associated with James O'Connor's (1973) influential *The Fiscal Crisis of the State*, which developed from within a Marxist intellectual tradition, and Richard Musgrave's important "Theories of Fiscal Crisis: An Essay in Fiscal Sociology" (Musgrave 1980), which largely followed Schumpeterian lines of analysis.

O'Connor observed that unproductive state 'social expenses', such as welfare payments, redress the 'contradictions of capitalism' and contended that tax systems under advanced capitalism are designed to effectively quarantine 'monopoly capital' from having to fund such unproductive state expenses. Under this analysis, the concurrent growth in monopoly capital and public sector outlays, and the associated privatization of surplus and socialization of 'social expenses', were presented as interrelated but internally contradictory, leading to a tendency towards fiscal crisis

in advanced capitalism. In contrast to Goldscheid, who assigned a primary role to public creditors in his analysis of exploitation, O'Connor emphasised the contradiction of advanced capitalism itself and the inability of the state to correct this contradiction. Consequently, instead of following Goldscheid prescription favoring the development of state capitalism, O'Connor advocated a reduction in the need for unproductive social expenses, probably though the replacement of advanced capitalism with a socialist system.

The response to O'Connor's work was mainly from authors writing in Marxist traditions, with critiques generally being favorable, although concerns were raised over the presentation of class struggle in the competitive and monopolistic sectors as incidental to market determined wage outcomes (Loxley 1982) and the difficulty of empirically testing O'Connor's thesis due to the abstract classification of public expenditures (Block 1981).

Richard Musgrave (1980) provided the definitive non-Marxist response to O'Connor, and to the general proposition that structural imbalances in the budget result in a general tendency to fiscal crisis. From a conceptual perspective, Musgrave emphasized that fiscal benefits and costs cut across class divisions and link better to group interests. For example, taxation activity is often stratified by income source, tastes or the pattern of household expenditure and age. Consequently, not only is the net incidence of fiscal activity too complex and multidirectional to be analysed in terms of the labour-capital dichotomy, but monopoly capital is also unlikely to be quarantined from the economic consequences of fiscal policy. From the empirical perspective, Musgrave

noted that the growth in social expenses (as a share of gross national product) has been negligible—which is also contrary to O'Connor's thesis. He also suggested that much of the observed increase in growth in other public expenditures could be accounted for by a particular factor; the growth in population from the post World War II baby boom.

### **Modern Theoretical Fiscal Sociology**

In addition to the perennial issue of fiscal crisis, other important issues to emerge in new theoretical approaches to fiscal sociologies include investigation of: sociological intolerances to tax policy; the relationship between macro and micro considerations; the methodological relationship between cameral sciences and fiscal sociology for studies of economic integration and regulation; the relevance of constitutional themes to fiscal sociology; and endogenous preferences where statecraft becomes soul-craft. The contemporary relevance of Paretian themes to the methodology and method of fiscal sociology is also under consideration.

More recently, the Schumpeterian approach to fiscal sociology was extended to investigate the sociological 'intolerances' associated with the tax structure (Christian Seidl's 1984). When public perceptions of the distribution of the tax burden do not reflect evolutionary change in social values, sociological intolerances begin to emerge over the tax system. Within this context, relevant considerations for tax policy include the evolution of social perceptions in regard to: the relative incidence of taxation on consumer and producer sovereignty; the extent of tax related income redistribution; the degree of privacy relative to the degree of avoidance associated with the tax



system; and the extent of non-political power enjoyed by private sector benefactors whose actions influence the need for taxation.

Recent investigations have emphasized the importance of micro considerations to fiscal policy, concluding that an enhanced understanding of the flexibility and subtlety of fiscal arrangements in mixed economies requires investigators to be cognoscente of both macro and micro issues (Musgrave 1991). The importance of this approach is that it suggests that Schumpeter cannot be followed in taking public expenditures as largely given or in ignoring public finance insights into factors such as 'externalities'. The inference here is that, when investigating the fiscal phenomenon, the insights of both public finance theory and fiscal sociology are complementary - a position that is consistent with the Italian approach to fiscal studies pioneered by Borgatta and Sensini. More recently, the methodological foundations of the European tradition in the cameral sciences and *Staatswissenschaften* have been proposed as a sound basis for casting fiscal sociology as a policy study, particularly for the investigation of issues in economic regulation and integration between jurisdictions, such as in the European Union (Backhaus 2002). In this context, fiscal sociology effectively provides the synthetic unification of economic, sociological, political and legal approaches to public policy issues.

Another promising and recent development in fiscal sociology is the attempt to introduce the principles of constitutional political economy to fiscal sociology (Wagner 2002). This approach contrasts the reliance of exogenous preferences in traditional public finance

with a reinvigorated fiscal sociology where statecraft and fiscal activity are recognized as preference altering activities. Within this framework, rules and procedures of human governance, including the rules governing the operation of fiscal institutions, play an important role in shaping preferences, with statecraft treated as a form of 'soul-craft'. Like the normative basis of constitutional political economy, the constitutional approach to fiscal sociology is directed towards establishing rules where constitutional frameworks support liberal governance. The constitutional approach to political economy, like its Schumpeterian predecessor, also has the capacity to consider the development of fiscal institutions and their impact on fiscal programs.

A new approach to fiscal sociology, inspired by Pareto and the Italian approach to fiscal studies is also now emerging (McLure 2003). It considers the relationship between the degree of decentralization in fiscal programs relative to the interacting forces for continuity or change in economic balances (between speculator and rentier activities) the political balances between (between indirect persuasion and direct confrontation) and balances in social behavior (between non-conformist and conformist conduct). A limitation of this generalized approach is that particular fiscal institutions and the actual 'form' of public programs become secondary to the degree of fiscal decentralisation on the one hand and the degree of decentralization in the economic, political and socio-behavioral balances on the other. Nevertheless, this limitation can be offset to some extent by considering the influence of rules and fiscal institutional on the actual form of fiscal programs when the

degree of fiscal decentralization is investigated.

In addition to the theoretical and methodological work undertaken as part of a modern renewal of interest fiscal sociology, there have also been a number of applied studies. These have focused primarily on the sociology of taxation

### **Applied Sociology of Taxation**

The sociology of taxation was a major theme in the early English language fiscal sociology of Fritz Karl Mann (1943). The issue has continued to be a subject of fiscal sociology. In this regard, John Campbell (1993) surveyed fiscal research literature to summarize the identified social attributes of taxation. The survey found that the main social determinants of taxation are broad, ranging from war, macroeconomic circumstances and crisis, to classes, interest groups and the centralization of the state. Similarly, the main identified consequences of taxation were also found to be broad, extending from political rebellion and state building to economic organization, labor force participation and philanthropy.

The most notable and recent applied fiscal sociology of taxation is the study of Japan's property tax revolt in the 1990s by Jinno and DeWit (1998). As well as focusing on Japan, the study reflects an applied approach that has general relevance to fiscal sociology. They employ an essentially inductive historical technique to establish the role that fiscal institutions play in the development of taxation policy. The emphasis is not on individualistic rational choice constrained by institutional rules, as is the case in constitutional fiscal sociology, but on fiscal institutions as autonomous entities with their own organizational identities. To facilitate

analysis, a series of implicit steps are followed.

Step 1: establish, through historical observation, the key defining elements of the fiscal system under consideration. In the case of Japan, Jinno and DeWit emphasize the aggressive re-centralization of taxing powers in the 1940s as a result of American reforms and the relatively decentralized distribution of public spending responsibilities. This fiscal system is therefore characterized as a 'centralized-deconcentrated system', with the power to tax being largely centralized while the responsibility for expenditure not being fully concentrated in the central government.

Step 2: establish, again through historical observation, the taxation policy agendas pursued by the key public sector bureaucracies. In the case of Japan, the Ministry for Finance is found to be a force for the increased centralization of taxation, forcing local governments to rely more on central government funding than on local taxation. In contrast, the Ministry for Home Affairs is a force that acts in opposition to tax centralization, because it has authority over local taxation.

Step 3: examine the incidence and outcome of institutional competition to influence the government's taxation policy agenda. In the case of Japan, there was a successful Ministry of Finance push to 'reform' taxes by permitting central government to access the property tax base. The Ministry of Finance effectively outflanked the Ministry of Home Affairs. It rationalized the proposal in terms of providing the central government with an instrument to constrain speculative investment in the Japanese property bubble. However, subsequent public comments suggest that the proposal was

driven by an agenda aimed at increasing the central government access to a new growth tax base that was previously dominated by local government.

Step 4: re-examine the fundamentals of the fiscal system in light of the new tax policy, after accounting for the impact of the measure on economy, the community and political parties. In the case of Japan, the property tax had to endure a short pre-bubble period followed by a post bubble period, and the consequent property tax revolt. Quick action by the Government, based on Ministry of Finance advice, to reduce land tax rates, shifted pressure back onto the Ministry of Home Affairs over the use of property and asset taxes to fund local government. In the long term, the outcome was effectively an increase in tax centralization, with an associated reduction in the relative authority of the Ministry for Home Affairs.

Application of this historical institutional approach shows that institutions with fiscal responsibilities may have long term 'memories', or policy agendas, which can play an important role in the development of fiscal policy. Furthermore, it shows that fiscal institutions are not immune from strong social reaction to policy initiatives, either directly or indirectly, following a change in economic circumstances.

The theoretical and applied aspects of fiscal sociology have also been recently blended in historical studies. In this regard, John Hobson (1997) investigated the relationship between taxation, the fiscal system and foreign policy, supported by empirical case studies of Britain, Russia and Germany. This approach suggests that the cost of militarism during the free trade period, between 1860 and 1979, represented a constant force on States to

increase their indirect taxes, culminating in higher tariffs and ultimately shifting states into a protectionist period, subsequent to 1879. This effectively highlights links between the State and the national and international economy, including the mode of taxation employed by the State and the social relations that shape taxation policy. In this context, the state is effectively presented as an institutional ensemble, with the authority of the state only deriving from the group elements that comprise the ensemble. Although Hobson's work is essentially historical, the forces he identifies are of some contemporary relevance.

### **Outlook for Fiscal Sociology**

Recent developments in this field remain formative, and in some instances provisional, with a received view on the methodological and analytical framework of the new fiscal sociologies yet to emerge. For the field to develop into a truly policy relevant body of theory, a clearer consensus on the scope of subject must first emerge. In particular, the sociological dimension to fiscal studies will need to be defined as something distinct from, and broader than, ad hoc notions of fiscal 'politics'. Nevertheless, applied sociologies of taxation are now reasonably well developed, with further research likely. Further research is also likely on the application of fiscal sociology to governance in developing countries, particularly in relation to rent seeking activities of oil rich developing countries Moore (2001) represents a first step in this direction.

### **Selected References**

Backhaus, J. (2002) "Fiscal Sociology: What For?" *American Journal of*

- Economics and Sociology*, Volume 61, Number 1, pp. 55-78.
- Block, F. (1981) "The Fiscal Crisis of the Capitalist State" *Annual Review of Sociology*, Volume 7, pp. 1-27.
- Borgatta, G. (1913) "I Problemi Fondamentali della Scienza Finanziaria", *Atti Reale Accademia delle Scienze di Torino*, Volume 48, pp. 985-1007
- Borgatta, G. (1920) "Lo Studio Scientifico dei Fenomeni Finanziari", *Giornale degli Economisti*, Volumelv, pp. 1-24, 81-116.
- Campbell, J. (1993) "The State and Fiscal Sociology", *Annual Review of Sociology*, Volume 19, pp.193-185.
- Goldsheid, R. (1925) "A Sociological Approach to Problems of Public Finance" ("Staat, öffentlicher Haushalt und Gesellschaft, Wesen und Aufgaben der Finanzwissenschaften vom Standpunkte der Soziologie") translated by Elizabeth Henderson, in Richard A. Musgrave and Alan T. Peacock (Editors), *Classics in the Theory of Public Finance*. London: Macmillan, 1967, pp.202-213.
- Griziotti, B. (1950) "Alla Scuola di Vilfredo Pareto e Maffeo Pantaleoni", in *Vilfredo Pareto: l'Economista e il Sociologo*, Università Bocconi (ed), Malfasi, Milano, pp.340-346.
- Jinno, N. and A. DeWit. (1998) "Japan's Taxing Bureaucrats: Fiscal Sociology and the Property-tax Revolt", *Social Science Japan Journal*, Volume 1, Issue 2, pp.233-246.
- Hobson, J. (1997) *The Wealth of States: A Comparative Sociology of International Economic and Political Change*, Cambridge University Press, Cambridge.
- Loxley (1982) "Fiscal Sociology and the Fiscal Crisis of the State—A Review of Goldscheid and O'Connor", *Contemporary Crisis*, Volume 6, pp. 285-301.
- Mann, F.K. (1943) "The Sociology of Taxation", *Review of Politics*, Volume 5, pp. 225-235.
- Mann, F.K. (1947) "The Fiscal Component of Revolution: An Essay in Fiscal Sociology", *Review of Politics*, Volume 9, pp. 331-349.
- McLure, M. (2003) *An Italian Foundation for New Fiscal Sociology: A Reflection on the Pareto-Griziotti and Pareto-Sensini Letters on Ricardian Equivalence and Fiscal Theory*. University of Western Australia, Economics Discussion Paper Number 03.11.
- Moore, M. (2001) "Political Underdevelopment: What Causes 'Bad' Governance" *Public Management Review*, Volume 3, Number 3, pp. 385-418.
- Musgrave, R. (1980) "Theories of Fiscal Crises: An Essay in Fiscal Sociology" in Henry J Aaron and Michael J Boskins (Editors), *The Economics of Taxation*. Washington DC: Brookings Institution, pp. 316-390.
- Musgrave, R. (1992). "Schumpeter's Crisis of the Tax State: An Essay in Fiscal Sociology", *Journal of Evolutionary Economics*, Volume 2, Number 2, pp. 23-47.
- O'Connor, J. (1973) *The Fiscal Crisis of the State*. New York: St. Martin's Press.
- Pareto, V. (1917) "A Benvenuto Griziotti", letter of 2 September 1917, in Giovanni Busino (Editor), *Correspondance, vol 19/2, Vilfredo Pareto Œuvres Complètes*. Geneva: Droz, 1975, pp.983-985.
- Schumpeter, J.A. (1918) "The Crisis of the Tax State" ["Die Krise des Steuerstaats"], in Alan T. Peacock (Editor), *International Economic Papers*, Number 4., MacMillan. 1954, pp.5-38.

- Seidl, C. (1984) “The Tax State in Crisis: Can Schumpeterian Public Finance Claim Modern Relevance”, in Christain Seidl (Editor), *Lectures on Schumpeterian Economics*. Berlin: Springer-Verlag, pp.89-109.
- Sensini, G. (1929) “Cenni di finanza teorica”, *Giornale degli Economisti*, Volume 44, pp. 373-384.
- Wagner, R.E. (2002) “States and the Crafting of Souls: Mind, Society and Fiscal Sociolgy”, in the proceedings volume from the *First Erfurt Conference on Fiscal Sociology*.

*Michael McLure*  
*Department of Economics*  
*University of Western Australia*  
*Perth, Australia*  
*michael.mclure@uwa.edu.au*

## Free Banking

*Kam Hon Chu*

### **Introduction**

Traditionally free banking refers to private banks' unrestricted competitive provision of currency. As financial innovations and banking evolution have blurred the definition of money, free banking in the modern literature refers broadly to application of *laissez-faire* principles to money and banking. It is sometimes used loosely as a synonym for total banking deregulation or financial liberalization.

Ideally, a free banking system should not have government barriers to entry or exit in the industry, restrictions other than the enforcement of valid contracts on issuing banknotes and deposits, central bank acting as an *ex ante* lender of last resort, government-run deposit insurance, legal reserve requirements, minimum capital requirements, restrictions on branching, banks' activities and the nature of financial services, and interest-rate controls (Sechrest 1993:3). These stringent criteria cast doubt on the existence of free banking in reality. Nonetheless, free banking as a theory serves to advance our understanding of government's role in the monetary system and its operation based on free-market principles. Free banking also has relevant implications for public policy issues like banking regulation, monetary policy, governance of both private and central banks, exchange-rate arrangements, etc. In fact, the renaissance of free banking partly arises from central banks' unsatisfactory performance in controlling inflation in the 1970s. Free banking is advocated as an alternative monetary system to maintain monetary stability.

Besides monetary reform, free banking has significant policy implications for financial system regulation and deregulation. A crucial question at the heart of the debate is: How should financial institutions be regulated and supervised? Globalisation and financial liberalization are often named as culprits causing economic and financial crises during the last two decades. Financial re-regulation is thus called for to promote stability. But many incidents, notably the US Savings and Loans debacle, suggest that government policies and interventions often do more harm than good in promoting financial stability and efficiency. To free banking proponents, the solution is not tighter regulation or re-regulation but more market discipline and freedom for all types of financial institutions.

### **Modern Free Banking Systems**

To facilitate the discussion of the pros and cons of free banking, we outline first the major forms of free banking in the modern literature. While *laissez-faire* is a common tenet, each approach has its own notions about the theory and practice of free banking.

Chronologically, the first approach to free banking is commonly known as currency competition (Klein 1974; Hayek 1978; Vaubel 1977; Engels 1985, among others). It does not imply abolishing governments' right to create money. Instead, it advocates that private banks, firms or even individuals be free to create their own identifiable currencies to compete with government monies, while citizens be free to choose their preferred currencies. In theory it is in the issuers' interests to maintain their reputation, i.e., their monies' purchasing power. Contrary

to Gresham's Law, currency competition and substitution means "good monies would drive out bad," leading to an optimal solution that best fulfills the private sector's needs.

The second approach is the Black-Fama-Hall (BFH) system, named after its originators (Black 1970; Fama 1980; Hall 1984) and also known as the "New Monetary Economics" (Cowen & Kroszner 1987). In this system the government determines on a non-coercive basis, or the market agrees upon, a common unit of account in terms of a certain quantity of a good or a bundle of commodities. The government's role is not to issue base money or to regulate banks but to ensure due enforcement of contracts. Meanwhile, private banks competitively issue distinctive inside money redeemable in an indexed quantity of the specified unit of account. The system is essentially a cashless economy and a pure accounting system of exchange, in which banks play the dual role of providing bookkeeping services and managing financial portfolios for depositors. The absence of base money would bring macroeconomic stability because of the disappearance of monetary disorders (see Greenfield & Yeager 1983, Yeager 1985, for details).

The third approach is unregulated competitive supply of inside money redeemable in a single distinct base or outside money, possibly but not necessarily a precious metal (White 1995; Selgin 1988). Though also known as the modern free banking school, it closely resembles the Free Banking School of the 19<sup>th</sup> century, which engaged in monetary debates with the banking and currency schools (Fetter 1965). Under this approach, profit-maximizing banks are allowed to issue any amount of inside money as long

as they can persuade the public to hold their liabilities. Competition would force banks to make their liabilities convertible into gold, or silver, or any real commodities upon noteholders' demand. The free market would determine the ideal convertibility contract. Competition would drive banks to hold optimal reserves against their liabilities even without any legal reserve requirements. The need to maintain convertibility and the law of reflux would force banks to limit the total amount of currency in circulation (see later and the Appendix for more detailed discussion).

### **The Case Against Free Banking**

The various arguments against free banking can be succinctly stated as: Free banking is unstable and inefficient because of market failures arising from factors like natural monopoly, asymmetric information, public goods, and externalities.

A conventional argument against free banking is that the issue of currency is a natural monopoly as the marginal cost of issuing banknotes is negligible, if not zero. This implies that a single bank is more efficient in supplying currency than any combination of small banks. Therefore, it is inconsistent for banking to be unregulated and competitive simultaneously (Laidler 1992).

Furthermore, there are economies of standardization when the market adopts a single unit of account. The uniformity of money is a public good and has externalities on the economy. Thus free banking would impose large information requirements on private transactors for they have to assess the quality of private monies, whereas the uniformity of money reduces this information burden. On the other hand, transaction externalities arise

because of transaction-cost savings to economic agents by using a common medium of exchange. All these externalities justify government's monopoly in the provision of money.

More importantly, these externalities due to money's natural monopoly and public-good characteristics make its private production socially suboptimal in a free market. If profit-maximizing producers cannot sell the external benefits, they would underproduce or not produce at all. Contrarily, without bearing all the costs, they would produce more than the socially optimal amount.

Another popular argument against free banking thus follows; free banking is more prone to monetary instability than central banking because of overissue under competitive supply of money and its undesirable macroeconomic consequences like hyperinflation and price volatilities that distort the optimal allocation of resources (Friedman 1960). In an extreme case, money would be worthless as a result of oversupply and the economy would ultimately return to barter and eradicate the vital role of money in facilitating exchanges and promoting growth (Pesek & Saving 1967).

Despite certain stabilizing mechanisms, such as the law of reflux (see later for details), that might check an individual bank's overissue, they would not function properly under the celebrated "in-concert overexpansion" hypothesis (e.g. McCulloch 1831 and recently Goodhart 1988)—when all banks expanded together in the same proportion, say, due to competition for market shares or collusion. Interbank adverse clearings would not check overissue because the reflux of each bank's increased notes would be exactly offset by other banks' expanded volume of

notes. Consequently, the expansive banks would not be disciplined by losses of specie reserves.

Under floating exchange rates, free banking would similarly render exchange-rate volatilities amid erosion of the public's confidence regarding the future purchasing power of a currency. Furthermore, a high degree of currency substitution would possibly increase the scope for speculative instability in exchange rates and corresponding price fluctuations (Woodford 1991). Under currency competition, there would be huge transaction costs of switching between currencies, information costs of following exchange-rate movements and exchange-rate risk. As free banks could not produce enough public confidence to maintain stable currencies, a central bank is needed to protect the purchasing power of a currency and to reduce the undesirable macroeconomic impacts of exchange-rate uncertainties. Some economists further call for Tobin-type taxes and capital controls as measures against destabilizing hot capital flows.

Besides exchange-rate and price instabilities, free banking would lead to more volatile business cycles than central banking. As Goodhart (1988:50) describes, "free banking tends at best to be procyclical and at worst may be directly responsible for severe fluctuations" because of competitive pressure. Unregulated banking would also reduce the effectiveness of monetary control (Tobin & Brainard 1963). Overall, the macroeconomic instabilities associated with free banking would undermine economic growth and full employment.

Likewise, banking instability is another popular argument against free banking. A traditional belief has that "free trade in



banking is synonymous with free trade in swindling" (Tooke 1840:205). Free banking is prone to fraud and counterfeit, because heterogeneous issue of individual banknotes could lead to counterfeit (Cagan 1963) whereas difficulties in enforcing banknote contracts could result in fraud or overissue (Friedman 1960).

The modern debate, however, focuses on banking instability due to bank failures and contagious runs. A dominant view has that the financial system is fragile (e.g. financial instability hypothesis; Minsky 1977) and hence a central bank is needed to promote stability because the production of confidence is a natural monopoly. The government can build public confidence more efficiently than private issuers because of certain fixed costs in building confidence and government's ability to tax (Klein & Melvin 1982). Also, centralization of reserve holdings by the central bank can achieve economies of scale (Goodhart 1988). Contrarily, more currencies imply higher transaction and information costs, and higher resource costs when individual banks hold reserves against their currencies.

More important, free banking is inherently unstable because of confidence externalities and the free-riding problem. The possibility of free riding means an individual bank, and hence the whole system, lacks incentives to promote sufficient confidence, say, by holding adequate reserves. When depositors lose confidence and trigger a redemption run, a lender of last resort is needed to curtail the run and to restore stability.

One factor leading to free riding and externalities is information asymmetry. Being a public good easily abused by free-riders, bank reputation can fail to deter contagious runs, because "it is extremely

difficult to distinguish between a relative high rate of return that is offered because of greater efficiency and one that is offered because the institution is also undertaking a much riskier strategy" (Goodhart 1988:64). Given asymmetric information, bank runs can be contagious because depositors, particularly small depositors who do not have the expertise to monitor banks or it is too costly for them to do so individually, cannot discriminate between "healthy" and "problem" banks. To stabilize the banking system and to protect depositors, banking regulation, deposit insurance and lender of last resort are needed. Similarly, Miles (1995) extends the asymmetric information argument to justify capital-adequacy requirements because of difficulties for depositors to assess banks' capital values.

The above traditional arguments against free banking stability are reinforced by financial globalization. Theoretically, internationalization of banking implies increased competition and risk exposures leading to higher risk of bank failures, whereas liberalization of capital flows increases volatilities of financial markets and fuels international transmissions of financial crises and contagions. Empirically, financial liberalization does increase the probability of a banking crisis (Demirguc-Kunt & Detragiache 2001). Therefore, an international lender of last resort is arguably needed to avert the risk of international banking crises.

### **The Case for Free Banking**

The case for free banking builds upon a simple question: If, in theory, perfect competition would lead to Pareto optimality in an industry, why should money and banking be an exception? If not, free banking would be both efficient and stable.

First of all, the public-good argument applies to money's unit-of-account function only. As a medium of exchange and a store of value, money is not a public good because it renders services only to its holders, i.e., non-rivalry and non-excludability in consumption. There is no divergence between the private and the social optimum in the choice of a common exchange medium. Historically, the emergence of gold and silver as money suggests the absence of transactions externalities, as the market would produce a commonly acceptable medium of exchange.

Neither is there an a priori case nor empirical evidence in favour of money and banking as a natural monopoly (Dowd 1992b). Even if it is, government regulations like barriers of entry to the banking industry are unnecessary, because the threat of potential entry would encourage the natural monopolist to operate efficiently like a competitive bank (Vaubel 1985). Although the marginal costs of issuing currency (the costs of paper and ink) are negligible and probably falling because of economies of scale, the costs of maintaining banknotes in circulation (e.g. marketing and liquidity costs) are definitely non-zero and rising under normal circumstances. The traditional overissue argument against free banking is hence simply a misconception about costs.

Furthermore, banks' needs to maintain brand names or reputation to keep their notes in circulation would prevent overissue under currency competition. In the Selgin-White system, the law of reflux would govern the amount of notes issued. Any excess supply of banknotes would rapidly return to their issuers through noteholders' direct redemption or adverse

clearings. Moreover, issuing banks themselves would monitor overissue by their rivalries through note duelling. The over-expansive bank would be disciplined by a loss of reserves and end up in a suboptimal position. Given positive marginal costs of maintaining banknotes in circulation and the above self-regulating mechanisms, overissue would not be possible in the long run. Consequently, free banking could maintain price and exchange-rate stabilities.

By the same token, given a fixed quantity of reserves "in-concert overexpansion" is unlikely to occur in reality and also self-correcting because the whole banking system would be checked by an external drain of specie through the Humean price-specie-flow mechanism (White 1995; Selgin 1988). In practice, banks would respond to profit signals and create credit and deposits demanded by the public. Over-expansion would increase banks' default risk and precautionary reserve demand (Selgin 2001). In sum, free banking is not inflationary and unstable. Contrarily, free banking would be more stable than central banking because there would no longer be multiple contractions or expansions of the money supply in response to changes in the base money. In fact, almost all inflationary episodes are due to central banks' monopolies in money creation, allowing overissue without any restraint. Furthermore, governments' incentives to extract seigniorage or inflation taxes may lead to political business cycles or seigniorage cycles (Wagner 1977). From this perspective free banking, like other monetary arrangements such as gold standard or monetary rules, serves to discipline governments from inflationary finance and to maintain monetary stability (White 1983).

Similarly, government interventions can be a source, rather than a cure, of financial instability. With short delays between the issue and reflux of private banknotes, clearings would check issuers' creditworthiness and reduce the chances of fraud. By contrast, a monopoly bank, if aided by a government-sanctioned suspension of payments, could defraud its clients and raise the danger of fraud (Selgin 1988).

Besides less prone to fraud, free banking would achieve stability and efficiency. Banking crises are simply not free-market phenomena (Selgin 1994). Without restrictions on note issues, banks could meet deposit withdrawals or runs by swapping banknotes for deposits, leaving their reserves unchanged. The necessity of a central bank for averting contagious runs due to confidence externalities and asymmetric information is also debatable. First, asymmetric information does not imply contagious runs, because in a separating equilibrium high-quality banks can signal their underlying quality, say, by offering lower deposit rates and holding smaller proportions of risky loans than low-quality banks (Chu 1999). Similarly, Dowd (1999) refutes the argument for capital regulation based on asymmetric information.

By the same token, banks have incentives to produce sufficient public confidence to avoid runs and to support any solvent rival suffering a contagious run. Certain stabilizing mechanisms to deal with bank runs would evolve in a free banking system. Historical examples include clearinghouse certificates and loan certificates (Timberlake 1984; Mullineaux 1987; Gorton & Mullineaux 1987) as a short-lived form of emergency high-powered money, option clauses in deposit

contracts which allow banks the option to defer note redemption so long as the noteholders are compensated (White 1995; Dowd 1988; Selgin 1988), and suspension of payments or convertibility as the welfare costs of suspension are likely to be less than those of government deposit insurance (Gorton 1985; Selgin 1993).

Hence, the conventional view that banking is inherently unstable is simply flawed. Government regulations, deposit insurance, and lender of last resort are not necessarily more effective and less costly than the market solutions. Even worse, central banks' assistance to poorly managed banks and failure to bail out problem banks in a timely manner lead to the notorious moral hazard problem that encourages bankers to take unwarranted risk resulting in higher probabilities of bank failure. Therefore, it is doubtful whether central banks actually create or destroy public confidence necessary for maintaining banking stability (Dowd 1989; Selgin 1989).

As on the centralization of reserves, the potential gains in economies of scale may be quite small as banks are already operating on low reserve ratios. Even if there are economies of scale in reserve holdings, a clearinghouse can take up the role in lieu of a central bank (Dowd 1989). Overall, the potential benefits of a central bank would not necessarily be large enough to outweigh the costs.

As free banking is not inherently unstable and does not reduce the efficiency of intermediation, there is no reason to believe why free banking would restrict growth and employment. The key question is whether the amount of lending and investment financed by a free banking system would be greater or less than that required for promoting maximum

sustainable economic growth (Selgin 1988).

As free trade promotes growth, there is no reason to believe why free banking—the free trade of financial services—should be an exception. Financial globalization would reduce the risk of an international banking crisis through international portfolio diversification. Moreover, bank governance and control would be strengthened through international bank mergers and acquisitions. The financial fragility associated with financial liberalization is due to many factors like macro-policy failures, ineffective regulation, absence of effective market discipline, etc. To ensure international financial stability, regulators should neither impose restrictions on the internationalization of the banking industry nor extend the international role of lender of last resort (Milne & Wood 2003).

### **Empirical Evidence**

Historically more than sixty free banking episodes existed in various forms (Schuler 1992). The well-known Scottish experience during 1716-1844 is often taken as evidence of successful practice of free banking (White 1995), mainly because of its lower bank failure rate and less frequent counterfeit of banknotes than its English counterparts (White 1995:ch.2; Fetter 1965:71-3). However, whether it was a genuine free banking system remains controversial, even among free banking proponents themselves (e.g. Goodhart 1988, Rothbard 1988, Sechrest 1988, to name just a few). The major debates include the lender-of-last resort roles of the Bank of England (Checkland 1975) and the two largest Scottish banks (Goodhart 1988) to other Scottish banks, the requirement of unlimited liabilities for all except three

large chartered banks as a barrier to entry (Carr et al 1989; Cowen & Kroszner 1989), and the Act of Parliament in 1765 which suppressed the issue of small notes and option clauses (Dowd 1989:122). See White (1995:ch.3) for a rebuttal of these arguments.

Another famous episode, the U.S. free banking era of 1837-1863, was commonly viewed as a disaster until a series of empirical studies, such as Rockoff (1974, 1975, 1985, 1991), King (1983), Rolnick and Weber (1983, 1984, 1988), and Economopoulos (1986, 1988, 1990), among many others, reveal that wildcat banking, contagious runs and note-holders' losses were exaggerated. Moreover, banknotes were priced properly by the market (Gorton 1996, 1997) and free banking generally increased competition and efficiency (Ng 1988). Falling asset prices were a primary reason for free bank failures. Like the Scottish case, whether the U.S. experience can be characterized as free banking is highly controversial, mainly because of the existence of numerous restrictions, such as requirement of government bond holdings against banknotes issued, usury laws, reserve requirements, capital requirements, restrictions on investment in real property, restrictions on branching, existence of deposit insurance in some states, etc (see Rockoff 1975 for details).

By documenting other countries' historical experiences, Dowd (1992a) concludes that free banking is not prone to inflation and banking instability. The prosperity of some countries during their free banking eras suggests promotion of economic growth by free banking (Cameron 1967; Sandberg 1978). Nonetheless, the empirical evidence remains mixed and controversial as some

recent historical studies suggest free banking instability in practice (e.g. Hickson & Turner 2002; Neldner 2003).

Not all free banking episodes are dated and took place before the 20<sup>th</sup> century. A recent example is Hong Kong before 1965, where banking regulation was lax and a central bank was absent. The experiences of the US, Canada and Hong Kong during 1935-64 show that free banking is not more prone to failures than regulated banking (Chu 1996). Hong Kong's experience not only suggests that in a separating signalling equilibrium asymmetric information does not imply instability due to contagious runs (Chu 1999) but also rejects the in-concert overexpansion hypothesis and that free banking is not cycle-amplifying (Chu 2002). However, the Hong Kong banking system was not entirely immune to banking crises, as it suffered in 1965 one of the most severe crises in its history.

It is not possible to discuss all the individual free banking cases here. However, it should be noted that none of them was a pure free banking system and was at best a close proxy for *laissez-faire* banking. Inevitably, the empirical findings and interpretations remain controversial. Nevertheless, these studies have undoubtedly contributed to our understanding of the operations of highly unregulated, if not totally free, banking systems.

### **Assessment of Free Banking**

We have selectively discussed the major arguments for and against free banking. Limited space does not permit in-depth analysis and discussion, let alone the omission of many theoretical and subtle issues such as the optimal quantity of money and the time inconsistency problem (for excellent surveys see Selgin 1988;

Selgin & White 1994). In brief, most of these arguments have existed in the literature for a long time. Despite the numerous studies in the last three decades, many theoretical issues are still unresolved while the empirical evidence is mixed and controversial. The mainstream view remains that free banking is infeasible in both theory and practice; and from the point of view of governance and public policy, central banks, regulations and safety nets are therefore necessary to promote financial stability and efficiency. By sharp contrast, free-banking theorists advocate that free banking is not only feasible but also more stable than central banking, as market discipline provides governance for both commercial and central banks.

The debate is further complicated by the trend of globalization over the last two decades. On the one hand, financial liberalization is commonly attributed to the frequent outbreaks of financial crisis, although the issue remains highly controversial. On the other hand, there is convincing empirical evidence showing that financial liberalization accelerates economic growth (Levine 2001). Thus, the risks and benefits of financial liberalization must be accurately assessed. Much work still needs to be done in order to create a "new financial architecture" to assure financial efficiency and stability without distorting the market mechanism.

To conclude, we should not believe in the omnipotence of the market because the possibility of market failure cannot be entirely ruled out. But at the same time, as the modern free-banking literature demonstrates, government failures are also possible. Since there are limits to both the market and government regulation, the study of free banking should not be ignored

in the search for an optimal public policy to promote financial stability and efficiency.

### Appendix: A Simple Mathematical Model of Free Banking Stability

As the BFH model and currency competition do not have historical precedents, we describe here only on stability of the Selgin-White model. Consider a representative free bank that chooses its levels of specie ( $S$ ), interest-earning assets ( $B$ ), banknotes ( $N$ ), and deposits ( $D$ ) to maximize its expected profit ( $\pi$ ) in a freely competitive system. Assume the bank's equity ( $K$ ) is exogenously given. Competitive markets imply that the yield rate on interest-earning assets ( $r_a$ ) and interest rate on deposits ( $r_d$ ) are also exogenously given. Mathematically, we have

$$(1) \text{MAX } \pi = r_a B - r_d D - C(S, B, N, D) - L(S, N, D)$$

subject to the balance-sheet constraint

$$(2) S + B = N + D + K$$

The last two RHS terms above are respectively the total operating cost function and the expected liquidity cost function, with the standard neoclassical assumptions of increasing marginal costs except  $L_S < 0$ .

The first-order conditions of the Lagrangean maximization problem are:

$$(3) r - C_B = -C_S - L_S = C_N + L_N = r_d + C_D + L_D.$$

For brevity, the familiar economic interpretations of these equimarginal conditions are not spelled out here (see White 1995 for details). The crucial result is: the profit-maximizing bank will not

expand its notes and deposits indefinitely because of rising marginal operating and liquidity costs.

Assuming a small open economy with fixed exchange rates on a specie standard, the whole banking system's equilibrium is characterized by the currency-specie ratio:

$$(4) \text{CUR}/S = (S_p + N_p^1 N_p^2 + \dots + N_p^m) / (S_p + S_1 + S_2 + \dots + S_m) \\ = (S_p/N_p + 1) / [S_p/N_p + S_1/N_p^1 (N_p^1/N_p) + \\ S_2/N_p^2 (N_p^2/N_p) + \dots + S_m/N_p^m (N_p^m/N_p)]$$

where  $CUR$  is the stock of currency held by the public (i.e., the sum of stock of specie held by the public ( $S_p$ ) and the stock of bank  $i$ 's notes held by the public ( $N_p^i$ )),  $S_i$  is the stock of specie reserves held by bank  $i$ ,  $S$  the monetary base and  $N_p$  is total banknotes held by the public.

So long as the public's desired ratio of specie to notes held and each bank's desired ratio of vault specie to notes in circulation both remain stable, the actual ratios tend to adjust towards the desired ratios through the adjustment mechanisms to restore equilibrium. Domestically, any excess notes would return to the issuing bank through direct redemption or through adverse clearings, causing a loss of specie by the expanding bank. If all banks expanded their notes simultaneously, the induced increase in imports would cause an outflow of specie, and the economy would adjust to restore its equilibrium through the Humean price-specie-flow mechanism.

Nonetheless, free banking stability remains controversial. For example, when the public loses confidence in the banking system its desired ratio of specie to notes may increase. Free banking, as Christ (1993) argues, is neither necessary nor sufficient for reducing cyclical instability.

## Selected References

- Black, F. (1970) "Banking and Interest Rates in a World without Money, The Effects of Uncontrolled Banking", *Journal of Banking Research* Volume 1, pp. 9-20.
- Cagan, P. (1963) "The First Fifty Years of the National Banking System", Chapter 2 in D. Carson (Editor) *Banking and Monetary Studies*. Homewood, Il: Irwin.
- Cameron, R. (1967) *Banking in the Early Stages of Industrialization*. New York: Oxford University Press.
- Carr, J., F. Mathewson and S. Glied (1989) "Unlimited Liability and Free Banking in Scotland", *Journal of Economic History* Volume 49, pp. 974-8.
- Checkland, S.G. (1975) *Scottish Banking, A History, 1695-1973*. Glasgow: Collins.
- Christ, C. (1993) "When is Free Banking More Stable Than Regulated Banking?" in H. Barkai, S. Fischer and N. Liviatan (Editors), *Monetary Theory and Thoughts: Essays in Honour of Don Patinkin*. London: Macmillan Press, pp. 185-209.
- Chu, K.H. (1996) "Is Free Banking More Prone to Bank Failures Than Regulated Banking?" *Cato Journal*, Volume 16, p.p 47-61.
- Chu, K.H. (1999) "Free Banking and Information Asymmetry", *Journal of Money, Credit and Banking*, Volume 31, pp. 748-62.
- Chu, K.H. (2002) "Do Free Banks Overexpand in Unison? Time Series Evidence from Hong Kong, 1954-66", *Kredit und Kapital*, Volume 35, pp. 1-18.
- Cowen, T., and R. Kroszner. (1987) "The Development of the New Monetary Economics", *Journal of Political Economy*, Volume 95, pp. 567-90.
- Cowen, T., and R. Kroszner. (1989) "Scottish Banking Before 1844, A Model for Laissez-faire?" *Journal of Money, Credit and Banking*. Volume 21, pp. 221-3.
- Demirguc-Kunt, A. and E. Detragiache. (2001) "Financial Liberalization and Financial Fragility", in G. Caprio, P. Honohan, and J. Stiglitz (Editors), *Financial Liberalization: How Fast, and How Far*. Cambridge: Cambridge University Press, pp. 96-122.
- Dowd, K. (1988) "Option Clauses and the Stability of a Laissez-Faire Monetary System", *Journal of Financial Services Research*, Volume 1, pp. 319-33.
- Dowd, K. (1989) *The State and the Monetary System*. Hertfordshire: Philip Alan.
- Dowd, K. (1992a) (Editor) *The Experience of Free Banking*. London: Routledge.
- Dowd, K. (1992b) "Is Banking a Natural Monopoly?" *Kyklos*, Volume 45, pp. 379-92.
- Dowd, K. (1999) "Does Asymmetric Information Justify Bank Capital Adequacy Regulation?" *Cato Journal*, Volume 19, pp. 39-47.
- Economopoulos, A. (1986) "The Impact of Reserve Requirements on Free Bank Failures", *Atlantic Economic Journal*, Volume 14, pp. 76-84.
- Economopoulos, A. (1988) "Illinois Free Banking Experience", *Journal of Money, Credit, and Banking*, Volume 20, pp. 249-64.
- Economopoulos, A. (1990) "Free Banking Failures in New York and Wisconsin, A Portfolio Analysis", *Exploration in Economic History*, Volume 27, pp. 421-41.
- Engels, W. (1985) "The Competitive Creation of Money, State-defined Currency and Free Issue of Banknotes", *Zeitschrift fur Wirtschafts—und*

- Sozialwissenschaften*, Volume 105, pp. 589-601.
- Fama, E. (1980) "Banking in the Theory of Finance", *Journal of Monetary Economics*, Volume 6, pp. 39-57.
- Fetter, F. (1965) *Development of British Monetary Orthodoxy 1797-1875*. Cambridge: Harvard University Press.
- Friedman, M. (1960) *A Program for Monetary Stability*. New York: Fordham University Press.
- Goodhart, C.A.E. (1988) *The Evolution of Central Banks, A Natural Development?* Cambridge, MA: MIT Press.
- Gorton, G. (1985) "Bank Suspension of Convertibility", *Journal of Monetary Economics*, Volume 15, pp. 177-93.
- Gorton, G. (1996) "Reputation Formation in Early Bank Note Markets", *Journal of Political Economy*, Volume 104, pp. 346-97.
- Gorton, G. (1997) "Pricing Free Bank Notes", *Journal of Monetary Economics*, Volume 44, pp. 33-64.
- Gorton, G., and D. Mullineaux. (1987) "The Joint Production of Confidence, Endogenous Regulation and Nineteenth Century Commercial Bank Clearinghouses", *Journal of Money, Credit and Banking*, Volume 19, pp. 457-67.
- Greenfield, R.L. and L.B. Yeager. (1983) "A Laissez-Faire Approach to Monetary Stability", *Journal of Money, Credit and Banking*, Volume 15, pp. 302-15.
- Hall, R.E. (1984) "A Free Market Policy to Stabilize the Purchasing Power of the Dollar", in Barry N. Siegel (Editor), *Money in Crisis*. San Francisco: The Pacific Institute.
- Hayek, F.A. (1978) *Denationalization of Money, The Argument Refined*. Second Edition. London: Institute of Economic Affairs, Hobart Paper 70.
- Hickson, C.R., and J.D. Turner (2002) "Free Banking Gone Awry: The Australian Banking Crisis of 1893", *Financial History Review*, Volume 9, pp. 147-67.
- King, R.G. (1983) "On the Economics of Private Money", *Journal of Monetary Economics*, Volume 12, pp. 127-58.
- Klein, B. (1974) "The Competitive Supply of Money", *Journal of Money, Credit and Banking*, Volume 6, pp. 423-54.
- Klein, B., and M. Melvin (1982) "Competing International Monies and International Monetary Arrangements", in M.B. Connolly (Editor), *The International Monetary System, Choices for the Future*. New York: Praeger.
- Laidler, D. (1992) "Free Banking, Theory", in J. Eatwell, M. Milgate and P. Newman (Editors), *The New Palgrave Dictionary of Money and Finance*. London: Macmillan, pp. 196-197.
- Levine, R. (2001) "International Financial Liberalization and Economic Growth", *Review of International Economics*, Volume 9, pp. 688-702.
- McCulloch, J.R. (1831) *Historical Sketch of the Bank of England: With an Examination of the Question as to the Prolongation of the Exclusive Privileges of the Establishment*. London: Longmans.
- Miles, D. (1995) "Optimal Regulation of Deposit Taking Financial Intermediaries", *European Economic Review*, Volume 39, pp. 1365-84.
- Milne, A., and G. Wood. (2003) "International Banking Crises", in A. Mullineux and V. Murinde (Editors), *Handbook of International Banking*, Aldershot, UK: Edward Elgar, pp. 403-27.
- Minsky, H.P. (1977) "A Theory of Systemic Fragility", in E.J. Altman and



- A.W. Sametz (Editors), *Financial Crises: Institutions and Markets in a Fragile Environment*. New York: Wiley, pp. 138-52,
- Mullineaux, D.J. (1987) "Competitive Monies and the Suffolk System: A Contractual Perspective", *Southern Economic Journal*, Volume 53, pp. 884-98.
- Neldner, M. (2003) "Competition Necessarily Tends to Produce Excess: The Experience of Free Banking in Switzerland", *German Economic Review* Volume 4, pp. 389-403.
- Ng, K. (1988) "Free Banking Laws and Barriers to Entry in Banking, 1838-1860", *Journal of Economic History*, Volume 48, pp. 877-89.
- Pesek, B.P., and T.R. Saving. (1967) *Money, Wealth and Economic Theory*. New York: Macmillan.
- Rockoff, H. (1974) "The Free Banking Era, A Reexamination", *Journal of Money, Credit and Banking*, Volume 6, pp. 141-67.
- Rockoff, H. (1975) *The Free Banking Era, A Reconsideration*. New York: Arno Press.
- Rockoff, H. (1985) "New Evidence on Free Banking in the United States", *American Economic Review*, Volume 75, pp. 886-9.
- Rockoff, H. (1991) "Lessons from the American Experience with Free Banking", in Forrest Capie and Geoffrey Woods (Editors), *Unregulated Banking: Chaos or Order*. New York: St. Martin's Press, pp. 73-109.
- Rolnick, A.J. and W.E. Weber. (1983) "New Evidence on the Free Banking Era", *American Economic Review*, Volume 73, pp. 1080-91.
- Rolnick, A.J. and W.E. Weber. (1984) "The Causes of Free Bank Failures", *Journal of Monetary Economics*, Volume 14, pp. 267-91.
- Rolnick, A.J., and W.E. Weber. (1988) "Explaining the Demand for Free Bank Notes", *Journal of Monetary Economics*, Volume 21, pp. 47-71.
- Rothbard, M. (1988) "The Myth of Free Banking in Scotland", *Review of Austrian Economics*, Volume 2, pp. 229-45.
- Sandberg, L.G. (1978) "Banking and Economic Growth in Sweden before World War I", *Journal of Economic History*, Volume 38, pp. 650-80.
- Schuler, K. (1992) "The World History of Free Banking", in K. Dowd (Editor), *The Experience of Free Banking*. London: Routledge.
- Sechrest, L.J. (1988) "White's Free Banking Thesis: A Case of Mistaken Identity", *Review of Austrian Economics*, Volume 2, pp. 247-57.
- Sechrest, L.J. (1993) *Free Banking: Theory, History, and a Laissez-Faire Model*. Westport: Quorum Books.
- Selgin, G. (1988) *The Theory of Free Banking: Money Supply Under Competitive Note Issue*. New Jersey: Rowman and Littlefield.
- Selgin, G. (1989) "Legal Restriction, Financial Weakening, and the Lender of Last Resort", *Hong Kong Economic Papers*, Volume 20, pp. 67-87.
- Selgin, G. (1993) "In Defense of Bank Suspension", *Journal of Financial Services Research*, Volume 7, pp. 347-64.
- Selgin, G. (1994) "Are Banking Crises a Free-Banking Phenomenon?" *Critical Review*, Volume 8, Number 4, pp. 591-608.
- Selgin, G. (2001) "In-Concert Overexpansion and the Precautionary Demand for Bank Reserves", *Journal of*

- Money, Credit, and Banking*, Volume 33, pp. 294-300.
- Selgin, G., and L.H. White. (1994) "How Would the Invisible Hand Handle Money?" *Journal of Economic Literature*, Volume 32, pp. 1718-49.
- Timberlake, R.H. (1984) "The Central Banking Role of Clearinghouse Association", *Journal of Money, Credit and Banking*, Volume 16, pp. 1-15.
- Tobin, J, and W. Brainard. (1963) "Financial Intermediation and the Effectiveness of Monetary Policy", *American Economic Review Papers and Proceedings*, Volume 53, pp. 383-400.
- Tooke, T. (1840) *A History of Prices and of the State of Circulation*, Vol. 3. London: Longman, Brown, Green, Longmans and Roberts.
- Vaubel, R. (1977) "Free Currency Competition", *Weltwirtschaftliches Archiv*, Volume 113, pp. 435-61.
- Vaubel, R. (1985) "Competing Currencies: The Case for Free Entry", *Zeitschrift fur Wirtschafts—und Sozialwissenschaften* 105, 547-64.
- Wagner, R.E. (1977) "Economic Manipulation for Political Profit: Macroeconomic Consequences and Constitutional Implications", *Kyklos*, Volume 30, pp. 395-410.
- White, L.H. (1983) "Fix or Float?", *Inquiry*, Volume 6, Number 12, reprinted in White L.H. (1989) *Competition and Currency: Essays on Free Banking and Money*. New York: NYU Press, ch. 7.
- White, L.H. (1995) *Free Banking in Britain: Theory, Experience and Debate 1800-1845*. Second Edition. Institute of Economic Affairs. (Originally published in 1984 by Cambridge University Press.)
- Woodford, M. (1991) "Currency Competition and the Transition to Monetary Union: Does Competition between Currencies Lead to Price Level and Exchange Rate Instability?", in A. Giovannini and C. Mayer (Editors), *European Financial Integration*. Cambridge: Cambridge University Press, 257-289.
- Yeager, L.B. (1985) "Deregulation and Monetary Reform", *American Economic Review*, Volume 75, pp. 103-7.
- \*\* I would like to thank the anonymous referees for their comments on an earlier draft. The usual caveat applies.

*Kam Hon Chu*

*Memorial University of Newfoundland  
St. John's, Newfoundland & Labrador  
Canada  
kchu@mun.ca*

## Free Trade and Protection: Theory and Perspectives

*Turan Subasat*

### **Introduction**

Free international trade is thought to be beneficial to all participating countries regardless of their country-specific circumstances as it leads to faster economic growth and promotes global income equality. This claim, however, is disputed by the fact that free trade is usually favoured by stronger economies. Most countries developed behind protective barriers, and allowed free trade only after they succeeded in developing international competitiveness (Chang 2005). Unsurprisingly, free trade theory emerged in the UK, the first industrialised country. In the 19<sup>th</sup> and early 20<sup>th</sup> centuries, continental Europe and the US did not echo the demand for freer trade. On the contrary, they were firmly protectionist. The US, the paragon of free trade, had by far the highest level of protectionist policies. It was not until the end of the Second World War, when the US became the strongest trading country, that its government advocated free trade. The then-weakened UK, together with the other European countries, remained protectionist until the 1960s.

This paper explores the theoretical literature and the theoretical aspects of empirical literature to reveal whether the recent liberalisation and internationalisation process can be rationalised by the alleged benefits of free international trade. Since it was introduced by David Ricardo as a justification of free trade, comparative advantage theory (CAT) has been modified and developed in various ways. The basic tenets of these theories, however, remained close to

Ricardo's original ideas. This paper critically reviews the various versions of CAT and explores their theoretical cogency. It then deals with the trade policy debate. Finally, the review of the empirical studies evaluates the existing literature to establish whether there is sufficient evidence in support of trade-liberalisation and export-promotion policies.

### **Comparative Advantage Theory**

In his *Wealth of Nations*, Smith (1776) argued that the major source of wealth was the extent of the division of labour and the international division of labour was the natural extension of the division of labour in domestic markets. Smith's argument was based on the absolute advantage theory (AAT), which asserted that commodities should be produced where the absolute cost of production is the lowest. Smith's AAT looked persuasive, but under this theory, trade was impossible when one country produced all commodities at a lower absolute cost than the other country. Ricardo (1817) dealt with this problem by developing the CAT which revealed that both countries could benefit from specialising in particular lines of production even where for each line of production, one country's absolute cost is lower than that of the other country. What was required was for relative productivities and prices to differ.

The following simple numerical example will clarify the gains from trade based on comparative advantage (CA). For the sake of simplification we have two countries (England and Portugal) which employ the same number of workers in the production of two commodities (wine and cloth). We assume constant returns to scale, full employment of all resources and balanced trade. Table 1 shows that Portugal

can produce more units of both commodities by employing the same number of workers, which implies that productivity levels in Portugal are higher. As prices are inversely related to productivity, both commodities are cheaper in Portugal than in England. According to AAT, there is no case for trade. Given lower prices in Portugal, England would benefit from importing both commodities from Portugal but this is not possible because Portugal would not benefit from trade as both commodities are more expensive in England.

Ricardo argued against AAT by demonstrating that if Portugal produces only wine and allows England to produce cloth, both countries would gain from trade, because total production would be higher. Portugal has CA in wine because it is relatively more productive in wine (productivity and price differences are larger between countries for wine) and England has CA in cloth because it is relatively less unproductive in cloth. With trade, England (Portugal) stops producing wine (cloth) and uses all the resources to produce cloth (wine). Therefore cloth (wine) production in England (Portugal) doubles. Although the total production of wine decreases by 10 units from 210 to 200, the total production of cloth increases by 50 units from 160 to 210. There is a clear gain for both countries if the net increase in production is evenly shared by both countries.

**Table 1. Comparative Advantage**

	Before Trade	Before Trade	After Trade	After Trade
	Cloth (units)	Wine (units)	Cloth (units)	Wine (units)
Portugal	100	110	200	0
England	60	100	0	200
Total Production	160	210	200	200

Some economists believe that CA is no more than stating that a country chooses to produce what it can produce best. In this sense, it implies neither free trade nor protection and does not involve any commitment to laissez-faire (see Schydrowsky 1984, Lal and Rajapatirana 1987). However, this is not how CA is generally understood in the literature. The theory has two important propositions which are both controversial: (1) Free trade amongst two countries is possible even when one county has no absolute advantage in terms of the cost of production; (2) a country will benefit from specialisation according to its CA regardless of the commodities involved. Although criticised heavily by some economists (see Shaikh 1999), the theory stands on stronger grounds with its first argument. The second argument however is weaker and needs to be considered in more detail.

Krugman (1998:35) argued that ‘Ricardo’s idea is truly, madly, deeply difficult. But it is also utterly true, immensely sophisticated—and extremely relevant to the modern world.’ There have been, however, various criticisms of Ricardo’s theory, from the Marxist left to the neoclassical right. The neoclassical economists accommodate Ricardian CAT, but reject its attachment to the labour theory of value. They have instead developed a ‘more sophisticated’ model of trade based on factor endowment theory that will be examined in the next section. Structuralist economists, such as Singer (1984), agree with the basic tenets of Ricardian theory and accept that international trade based on CA would indeed benefit all participating countries. They argue, however, that these benefits are unequally distributed between

developed and developing countries as a result of the nature of the commodities they export. They recommend temporary protectionist measures for the industrial sector, until international competitiveness in manufactured commodities is achieved. A stronger critique of CAT has come from the dependency theorists, particularly Emmanuel (1972) who argued that developing countries are exploited at the level of exchange because trade is based on 'unequal exchange'. This unequal exchange, he argued, occurs as a result of the wage and price level differentials between developing and developed countries.

Most criticisms of Ricardo's theory are levelled at the restrictive assumptions of the model. Hudson (1992) summarised these assumptions and criticised them in detail. The theory indeed runs into difficulties as a result of some of its unrealistic assumptions, but the basic arguments of the theory, which aim to justify trade based on comparative cost, survive even when the assumptions are relaxed. For example, probably the most frequently criticised assumption of the theory is the assumption of full employment which guarantees that there is no employment loss resulting from the introduction of free trade. Labour and capital are assumed to be fully mobile between the sectors and can be shifted from one sector to the others without causing unemployment which is clearly unrealistic. Structural inflexibilities may cause long-term unemployment of factors and may eliminate the benefits of trade. This is a valid critique of rapid trade-liberalisation policies but not of gradual liberalisation policies which would minimise the cost of adjustment and would still allow specialisation based on CA. It

may indeed be very difficult to shift resources from the production of one commodity to another. However, this transition from more diversified to more specialised production can be a gradual process rather than a rapid one. A country may direct the new resources into the production of commodities for which it has a CA rather than remove resources from one sector to another and reduce the transition cost. (For a more detailed discussion of the assumptions of the CAT, see Subasat 2002)

Three fundamental problems undermine the theory. The first important problem in Ricardo's theory concerns with how the static efficiency gains from free trade are distributed between countries. The model simply assumes that both countries benefit from the specialization in specific commodities as total production increases but fails to explain how these benefits are distributed between countries. In the model, trade disequilibrium is rectified by relative price changes. In other words, excess demand (supply) for a commodity would be eliminated by a rise (fall) in its price. This implies that the benefits of trade will depend upon the terms of trade which is not pre-determined but formed by a number of factors such as demand and supply shocks, the nature of tradable commodities and trade policies. Because it is possible for one country to capture most of the benefits, international trade is inherently an unstable process exhibiting fundamental conflicts between countries.

The second important problem of the theory is related to its static nature. Ricardo's theory is based on short-term static-efficiency gains through specialisation and 'implies that nations should live permanently in the short-run rather than maximising their productive

power over time' (Hudson 1992:116). When there is free trade, developing countries may indeed specialise in commodities for which they have CA as the theory predicts. This is precisely why they should avoid adopting free trade policies that may not be to their advantage in the long run. The 'dynamic' versions of CAT which attempt to address this problem will be considered in another section.

The third problem with the theory relates to its simplistic treatment of labour theory of value. The theory, at the highest level of abstraction, assumes that commodities are exchanged according to their values. In turn, the value of a commodity is determined by the necessary labour time employed to produce it. For example, if it takes two hours to produce one unit of wine and it takes one hour to produce one unit of cloth, then one unit of wine will be exchanged for two units of cloth. This is, however, an abstraction as two different types of labour power are considered as one for simplification purposes. It is simply assumed that one labour hour creates an equal amount of value in wine and in cloth production. In reality, however, the different types of labour power have different characteristics.

Marx recognises this and explains it in terms of 'complex' and 'simple' labour. According to Marx (1990:135), '[m]ore complex labour counts only as *intensified*, or rather *multiplied* simple labour, so that a smaller quantity of complex labour is considered equal to a larger quantity of simple labour.' For example, the value created by spending one labour hour in computer engineering is not equal to the value created by spending one hour in shoe polishing. In effect there are high skill, high technology and high value-added jobs

as well as low skill, low technology and low value-added jobs. Some jobs bring higher income simply because some types of labour power create higher value. This argument implies that specialisation according to CA will not benefit all countries. It is simply a mistake to specialise in low-skill, low-technology and low-income commodities, or commodities for which there is no future demand. Countries that specialise in such commodities will suffer. Consequently, what matters is not specialisation itself but the commodities in which countries specialise. This is not exploitation through unequal exchange but simply wrong specialisation resulting from free trade and CA (Subasat 2002).

It is advantageous for countries to specialise in commodities that bring long-term benefits. Since such commodities are always 'limited', conflict is a permanent feature of international markets. Countries compete over such commodities and continuously reposition themselves to achieve market domination. Such an approach to trade explains why free trade has always been a contentious issue and justifies protectionist and promotionist policies permanently if international competitiveness in the most beneficial commodities cannot be established. For an individual, it is probably better to be the worst off doctor in business than the best earning street sweeper. In the same manner, for a country, it is probably better to be a less efficient producer of a high-value added commodity than a more efficient producer of a low-value added commodity.

### **Heckscher-Ohlin Model**

For Ricardo, mutually beneficial trade between two countries was possible

because of the differences in pre-trade relative commodity prices resulting from differences in labour productivity. Neoclassical critiques of Ricardo, however, argued that he failed to explain the reasons for different productivity levels in different countries, and incorrectly assumed that labour was the only factor of production. The Heckscher-Ohlin (H-O) theory is a direct application and natural extension of the neoclassical analysis to international economics. It seeks to explain not only the reasons for the differences in the relative commodity prices by the 'factor endowments', but also the impact of international trade on these endowments. The theory was later 'refined' by Stolper and Samuelson (1941) who argued that free trade would also equalize the factor prices.

The H-O model is based on a number of very controversial assumptions: all countries use the same technology in production; factors of production are mobile domestically but immobile internationally; tastes are the same in all countries; there are no scale economies; there is perfect competition in all markets; all resources are fully employed. Given these assumptions, the theory asserts that a nation will export the commodities that make intensive use of the country's relatively abundant and cheap factors, and import the commodities whose production require the intensive use of relatively scarce and expensive factors. This means that in labour abundant (developing) countries, where the wage rates are low, labour-intensive goods can be produced for export relatively cheaply. Similarly, these countries will import capital-intensive commodities to compensate for their own scarcity of capital that renders capital-intensive production expensive, again in

relative terms. To determine if a country is capital or labour endowed (or abundant) we need to look at their comparative availability, namely capital-labour ratios. If one country has a higher capital-labour ratio than another, that country is endowed with capital, or is capital-abundant.

Stolper and Samuelson (1941) argued that if the H-O theory was accurate, then when international trade takes place, the prices of the factors would gradually converge until they become equalized in both countries. As one country specializes in the production of labour (capital) intensive commodities and reduces its production of capital (labour) intensive commodities, the relative demand for labour (capital) and the wage rate (interest rate) rises, and the demand for capital (labour) and the interest rate (wage rate) falls. Trade possibilities are exhausted when relative and absolute prices of the factors are equalized between countries.

The Heckscher-Ohlin-Stolper-Samuelson (HOSS) theory makes two predictions. First, most trade will take place between labour-abundant low-income countries and capital-abundant high-income countries, as trade possibilities are greater between countries that have dissimilar factor endowments. Second, as free trade takes place and countries specialise in particular commodities according to their factor endowments, the factor prices will be equalised internally as well as internationally. These predictions are not supported by historical evidence. Firstly, Leontief (1954) found that the US, as the most capital abundant country, exported labour-intensive commodities and imported capital-intensive commodities which contradicted the theory. Secondly, in contrast to the expectations of the theory,

both developed and developing countries have increased trade amongst rather than between themselves since the 1960s. Thirdly, there is no evidence of factor-price equalisation (Wood:1994). Empirically, the factor endowment theory has been rejected repeatedly. Trefler (1995:1029) suggests that 'the theory correctly predicts the direction of factor service trade about 50 percent of the time, a success rate that is matched by a coin toss.'

A number of critiques stay within the logical limits of the H-O theory and reflect the difficulties that the theory faces. For example, 'factor intensity reversal' refers to the situation where the same commodity (e.g. rice) is considered labour-intensive in a labour-abundant country (i.e. Vietnam), and capital-intensive in a capital-abundant country (i.e. USA). Another difficulty arises because capital is not homogeneous and cannot be aggregated. The H-O theory argues that if capital is abundant it will be cheap. One way to determine whether capital is abundant or not is to look at the capital-labour ratio. Since capital is not homogeneous, it can only be aggregated in term of its value which is a problematical exercise as the market value of capital cannot be calculated independently of the rate of interest (return on capital). If the interest rate increases, the cost of borrowing and the cost of capital also increase. Therefore, the whole theory reveals a chicken and egg dilemma (Edwards 1985). There is no unambiguous way to know if a country is labour- or capital-abundant since the value of capital cannot be determined independent of factor prices (interest and wage rates).

More importantly, the treatment of capital in the H-O theory is simplistic. Capital is considered as a non-produced endowment such as land, natural resources

and population. Its scarcity (abundance), then, determines its relative price. This is a very strong assumption that requires justification which the theory fails to provide. The only possible explanation for such an invalid assumption is that the production of capital takes time and in the short-run it can be considered as externally given to the economy. But what casts doubt upon this assumption is that while capital does not have the *time* to adjust, all the other factors do and this adjustment takes place simultaneously. The theory is therefore based on two assumptions that are hard to justify: firstly, that capital cannot adjust and secondly, that everything else adjusts simultaneously. For example, when free trade is allowed, countries have enough time to move resources from the production of one commodity to another, but they do not have enough time to produce more capital. This arbitrary handling of the concept of time reveals the theory's ideological mission in which 'short-run' is a mental construction to achieve desirable outcomes.

When capital is allowed to be produced, (assuming all countries have access to the same technology) there is no reason why one country should be endowed with more capital than another, and consequently there is no case for trade based on capital endowment. This is so, because if capital is produced there is no such thing as capital-scarcity or abundance. When there is 'capital-scarcity', the price of capital will temporarily rise until more capital can be produced. Moreover, once the clearly unrealistic assumption of the international immobility of factors of production is relaxed, capital-scarcity has no real meaning as capital can easily be imported. It is unreasonable to argue that



consumption goods, and not capital goods, can be traded.

Development requires capital accumulation and history demonstrates that nations are not endowed with capital but they accumulate it by deliberate policies (Hudson 1992). By failing to recognise the link between capital accumulation and development, the theory advises developing countries to focus on the production of labour-intensive, low skill, low technology and low value-added commodities. The reasons why only developed countries happen to be capital-abundant and developing countries capital-scarce, the most important question in development economics, remain unexplained. (See Subasat 2003 for a full discussion of the HOS model).

### **Dynamic Comparative Advantage**

Recognising the limits of static CAT has led to a wide range of studies attempting to develop a dynamic version of the theory. These versions seek to remove the static nature of typical CAT by focusing on the future rather than the current comparative cost structure. This section will argue that not all those theories are truly dynamic and the truly dynamic trade theories should not be called CAT.

There are a number of diverse arguments that can be grouped under the heading of dynamic CAT but due to space limitations, this section will only cover the infant industry argument which is the best-known form of dynamic CAT. Based on the assumption of increasing returns to scale and learning-by-doing, the infant industry argument provides the strongest rationale for short-term protectionism. It argues that developing countries should protect their industries from international competition and allow them to exploit

scale economies until they become strong enough to compete in international markets. As the initial cost of production and prices will be higher than those on the international markets, there will be an initial cost that a country must bear. If, however, average cost eventually falls lower than average international cost, the country will benefit in the long-run. Before deciding whether an infant industry should be protected, it is important to know how long it will take for an infant industry to mature. If it takes too long, the cost might be higher than the discounted future benefits. An infant industry should only be protected if there is scope for positive net benefits.

The dynamic versions of CAT denote a significant advancement over the static versions, since they recognise the possible changes in long-run comparative costs and the necessity of state intervention to facilitate structural change and development. 'Dynamic CAT', however, is a contradiction in terms. CAT is a static theory by its nature and cannot be made dynamic as it relies on static parameters such as the short or long-run production function and factor endowments that cannot be altered. CA refers to specialisation based on static (short or long-run) relative costs which condition the options of policymakers and guide them in the production of a particular bundle of commodities. The idea behind the dynamic CAT is that today's CA may not reflect true CA due, for example, to latecomers' disadvantages. Given its factor endowments, a country may have CA in a particular bundle of commodities but because the sector is in its infancy, it may not be able to compete with already established competitors. Therefore, a short-term protection is necessary for the sector

to utilize scale economies and learning-by-doing to allow long term real CA to prevail. Before production takes place, those commodities that have potential for future CA must be identified based on the static long-run production function.

In this view, protection is justified only temporarily until the industry (or a firm) 'matures' and only if protected sectors have potential CA in the future. Otherwise the market should be liberalised. The argument is not against the basic tenets of CA, but revolves around which commodities will have long-term CA as opposed to the current CA. Therefore this approach continues to carry the static characteristics of the conventional CAT. To put it another way, this version of the theory is not about creating competitiveness in the most beneficial (high value added) commodities by deliberate policies, but rather it is about accepting passively the possibility of low future comparative cost and protecting such industries until they grow up and become competitive in international markets.

Even in the dynamic versions of the theory, a country may have CA in low value-added commodities. The logic of the theory suggests that if the 'true CA' of a country is the same as the current CA, then the country should continue to specialise in currently produced commodities. As argued earlier, however, countries must avoid specialisation in low skill, low productivity and low value-added commodities, or commodities that have a very low income/demand elasticity and no future demand, whether or not they have short- or long-term CA. Dynamic CAT recognises that the current relative cost may not be a good indicator of the real relative costs, but does not question the

logic of the theory itself. Once the alleged benefits of static CAT are rejected, the arguments for the benefits of dynamic CAT also become doubtful.

As noted earlier, not all the dynamic theories suffer from the above weaknesses. They recognize the need to protect the domestic economy for a relatively long period of time until international competitiveness in the most beneficial commodities is established. Such a trade strategy requires careful planning and active involvement of the policy makers. When the 'endowments' and competitive advantages are created dynamically by the conscious decisions of the policymaker alone, however, this should not be called CA as CA refers to specialisation based on static (short- or long-run) relative costs which condition the options of policymakers. It is a concept that guides the policymaker in the production of a particular bundle of commodities. In a dynamic world, there may be no apparent short or long-run 'CA' in the production of a particular commodity but the conscious decisions taken today may bring future competitiveness through a combined process of protection and promotion.

Because future competitiveness is solely determined by the decisions taken today, and because the decision is not based on static short or long-run production function, such a dynamic approach cannot be called CA. In this sense, the success of a country in international markets is determined by its ability to develop competitiveness by reducing cost through a combination of successful industrial policy, experience, ability to produce and develop technology and good luck (in the sense that someone else is not more successful). Thus, competitiveness is not based on any short

or long-run CA at the beginning, but instead, is based purely on the success of policymaking. Arguably, the only CA a country has is a successful industrial policy. As Hudson (1992:31) argues '[h]istorically the first endowment required for industrialisation has been a policy of protectionism.'

### **Trade Policy**

Discussions about import-substitution and export-promotion as trade policies embody two related, but distinct, issues. The first is whether import-substitution and export-promotion policies are substitutes or are complements. The second is whether export-promotion is necessarily associated with trade-liberalisation and a reduction in the role of the state.

Import-substitution strategies aim to promote production by substituting domestic goods for previously imported consumer goods with the help of a wide range of tariff and non-tariff barriers and exchange rate policies. Looking at it from a short-term, static-efficiency point of view, neoclassical economists believe that import-substitution policies, by creating barriers to free trade and preventing specialisation, reduce incentives of the local firms to cut costs and increase productivity; lead to monopolistic structures; distort prices and misallocate resources.

The concept of export-promotion is more controversial than it at first appears. For some, export-promotion requires a neutral strategy with no bias against exports (Balassa 1989). According to others, however, export-promotion refers to policies which promote exports and does not suggest neutrality (Krueger 1985). In the first perspective, import-substitution is seen to involve inward-looking policies

with state intervention, while export-promotion focuses on outward-looking policies without state intervention. The question then becomes one of simply determining whether or not liberal and open economies grow faster. From this perspective, export-promotion policies beyond neutrality are seen as ineffective as import-substitution since they also distort prices. In the standard neoclassical two-sector production-function type framework, resources can be employed to produce either exportables or importables but not both. Therefore, import-substitution and export-promotion are necessarily substitutes and that export-promotion necessarily requires trade-liberalisation (Bhagwati 1988). To increase the production of exportables, the production of importables will have to decrease, and import controls will need to be liberalised (Liang 1992)

The above approach is too simplistic to deal with the complex realities of developing countries. To call neutrality of incentives as export-promotion is nonsensical since neutrality implies that neither sector is protected or promoted. The import-substitution and export-promotion dichotomy derives from the static, two-sector neoclassical model. In a more dynamic model where the static parameters of neoclassical economics are allowed to change, however, import-substitution strategies are not considered as the opposites of export-promotion strategies. Firstly, import-substitution is seen as a precondition for export-promotion, and secondly, import-substitution and export-promotion can be implemented together. Therefore, import-substitution is sometimes called 'import protection as export-promotion' (Krugman 1984).

Grabowski (1994) argues that both learning-by-doing and scale economies depend on the size of the market and both import-substitution and export-promotion policies guarantee the size of the market in turn. Initially, import-substitution policies allow demand for selected industries to grow faster than domestic consumption by protecting the domestic market and replacing imports by domestic production. When the limits of the domestic market are reached, however, export-promotion policies allow demand to grow faster than domestic consumption. Both policies, therefore, facilitate faster productivity growth due to greater investment, increasing returns to scale and learning by doing.

Liang (1992) shows that the production of exportables does not necessarily require a reduction in the production of importables once the two-sector model is replaced with a more realistic three-sector model. In a model where the economy is divided into three sectors (exportables, importables and non-tradables), not only is it possible to promote both import-substitution and export-promotion, but it is also possible to have anti-import-substitution and anti-export-promotion policies which would benefit the non-tradables sector.

The second perspective which goes beyond neutrality and defines export-promotion as a set of policies that leaves relative domestic rewards for exporting compared to importing greater than the rewards that would exist under free trade (Kruger 1985). This approach is conceptually more accurate but contradicts the basic principles of the neoclassical approach. According to neoclassical trade theory export-promotion is as inefficient as import-substitution because market prices

are distorted. Indeed, as Streeten (1982:162) argues, in its static form '[i]t is just as possible to have inefficient export policies as it is to have inefficient import-substitution.' Moving from an import-substitution strategy to an export-promotion strategy may eliminate some distortions but it may also create others.

The above arguments simply mean that export-promotion as well as import-substitution strategies make sense when the short-term, static approach of the neoclassical model is abandoned and a long-term dynamic perspective is adopted. Once this important point is clarified, export-promotion strategy, like import-substitution strategy, is incompatible with the neoclassical approach. As non-free market (interventionist) policies, their relationship should be analysed from a more dynamic perspective. Since both policies require state intervention, they are heterodox in their nature.

The rapid economic development of a number of East Asian economies, such as Japan, South Korea and Taiwan, provides overwhelming support for these arguments. Their success was initially attributed to liberal trade policies ('export-promotion') that they allegedly adopted. Haberler (1987), for example, argued that their success is fully explained by, and confirms, the neoclassical paradigm. Subsequent research, however, revealed that neoclassical economists were overly optimistic about the validity of their arguments in explaining this 'miracle'. As Chang (2005:49) argues '[d]espite some lingering disagreements, there is now a broad consensus that the spectacular growth of these countries, with the exception of Hong Kong, is fundamentally due to activist industrial, trade and technology policies by the state.' Rather

than relaying on trade-liberalisation to promote exports, these countries adopted a pragmatic combination of protectionist and promotionist policies as part of a broader industrial policy.

### **Theoretical Problems of Empirical Studies**

There is no consensus regarding the impact of alternative trade policies on economic development. Despite the prevalent view, which supports liberal trade policies, the empirical evidence is mixed. This section will not cover the empirical literature but highlight some of the theoretical complications that make the empirical assessment of alternative trade policies difficult.

Rather than seeking to measure trade orientation, earlier empirical work often attempted to estimate the statistical relationship between exports and economic growth rates to determine whether they were correlated. Any correlation found was interpreted as evidence of the benefits of promoting exports. This method was criticised, '[s]ince exports are themselves part of the national product, an autocorrelation is present; and a positive correlation of the two variables is almost inevitable, whatever their true relationship to each other' (Michael 1977:50).

Another important facet in the relationship between exports and economic growth rates is causality. A strong correlation proves neither the existence of causality between the two variables, nor if there is causality, does it mean causality runs from exports to economic growth. Most empirical studies, explicitly or implicitly, assume that causality runs from exports to economic growth. Jung and Marshall (1985) contend that even if it is true that export growth can cause economic

growth, it is equally plausible that economic growth may in turn cause export growth. For example, in the case of unbalanced growth, it is highly unlikely that the domestic demand for goods from expanding industries will grow as rapidly as their production. Therefore producers will be forced to seek out foreign markets in which to sell their commodities. In this case the causality will be from output growth to export growth. The 'stage of development' theory of CA argues that economic growth and successful exports performance are both determined by processes of development and structural change (Yaghmanian, 1994). As the economy becomes stronger, markets become more efficient and fewer bottlenecks occur. This well-functioning economy facilitates greater penetration into world markets (Pack 1992). As countries become more developed, they are more likely to 'get the prices right', and in so doing to follow a more neutral policy stance both with respect to exports and the domestic economy.

The failure of the earlier studies that used actual trade figures led to the creation of more complex measures of trade orientation that have their own weaknesses. Pritchett (1996) identified 'six trade policy stance measures': structurally adjusted trade intensity (SATI); Leamer's Openness Index; average tariffs; non-tariff barriers frequency; price distortion; and Leamer's trade distortion indices. In addition to these six measures of openness, different studies adopted a number of other measures to determine openness, such as the effective exchange rate ratio between exports and imports, real exchange rate devaluations, parallel market premium, effective rate of protection, direct controls such as quotas and import licensing schemes, export

incentives; degree of exchange rate overvaluation and a state monopoly on major exports.

All these measures have important deficiencies. Some measures have theoretical weaknesses and produce misleading results. Others are theoretically more accurate but difficult to calculate, particularly for a long period of time and across countries. They are non-comparable across countries and it is not possible to see a trend through time. Consequently, they cannot be used for time-series analysis. An evaluation of these measures cannot be done here because of limited space. Rodrigues and Rodrik (2000) provided a powerful critique of the above measures and the general approach to studying the relationship between trade policy and growth. They argued that

“[the] methodological problems with the empirical strategies employed in this literature leave the results open to diverse interpretations. In many cases, the indicators of ‘openness’ used by researchers are poor measures of trade barriers or are highly correlated with other sources of bad economic performance. In other cases, the methods used to ascertain the link between trade policy and growth have serious shortcomings.” (Rodrigues & Rodrik 2000:1)

Pritchett (1996) casts more doubt upon the accuracy of these measures by demonstrating that even though many studies that utilised different measures of trade openness demonstrate a positive correlation with economic growth rates, these measures themselves are uncorrelated. The lack of correlation may result from the fact that different measures attempt to gauge either export-promotion or trade-liberalisation. Since they measure

different things, it is not surprising that they are uncorrelated. If these different measures were all ‘accurate’ indicators of trade-liberalisation and export-promotion, one would expect them to be correlated at least within these two broad categories, i.e. the measures that intended to measure trade-liberalisation (export-promotion) would be correlated. Given that different measures of trade-liberalisation (export-promotion) are uncorrelated, only one of them, and probably none, could be considered ‘accurate’.

As argued earlier, export-promotion needs to be separated from trade-liberalisation, and once this separation is made, the relationship among export-promotion, trade-liberalisation and economic growth requires further elaboration. Measuring the impact of export-promotion on economic performance is a valid exercise. The same cannot be said for trade-liberalisation since the impact of trade-liberalisation on economic performance is normally transmitted through its impact on trade openness. Therefore, the questions become whether export-promotion leads to faster economic growth and whether trade-liberalisation promotes exports.

These two questions should be answered separately. This separation is essential because even if export-promotion was to lead to faster economic growth, trade-liberalisation may not be the best way to promote exports. The trade-growth link has been extensively investigated in the literature. The impact of trade-liberalisation on trade openness, however, is often assumed rather than proven. In this regard the recent limited literature produced mixed results. Although Santos-Paulino and (2004) found a positive link between trade-liberalization and export

growth, Subasat (forthcoming) found a very weak link between trade-liberalization and trade openness.

## Conclusion

This paper argues that the theoretical arguments and empirical evidence in support of trade-liberalisation is weak, and import-substitution and export-promotion policies should be seen as complementary rather than contradictory policies. The contention has not been made against the possible benefits of trade, but against the unsubstantiated belief in export-led development and trade-liberalisation. Behind the elegant theories of free trade lies a basic fallacy. Free trade and specialisation bring benefits to some but not all countries. The liberalisation process does not produce a natural tendency towards mutual benefit. It may create further disequilibrium and conflict. As a result, the 'existing advantages are reinforced and the resulting spatial distribution of economic activity is likely to exhibit strong divergence, leading to increasing inequalities within and between regions' (Kozul-Wright, 1995: 138). Free trade policies accelerate and reinforce uneven development on a larger scale. The world is disorderly and fractured, and the observed reality is uneven development.

From the perspective of developing countries, international trade is neither inherently good nor bad. Trade has the power to create opportunities and it has the power to destroy livelihoods. Developing countries have diverse structures. Industrial policies in general and trade policies in particular should be produced according to a country's specific circumstances. A uniform trade policy cannot bring about the benefits that it purports. Developing

countries should have a pragmatic approach to trade policies.

## Internet Sites

OECD. (See 'Trade') [www.oecd.org](http://www.oecd.org)

Oxfam International (Fair Trade) [www.maketradefair.com/en/index.htm](http://www.maketradefair.com/en/index.htm)

UNCTAD. [www.unctad.org](http://www.unctad.org)

World Bank. (See 'Trade') <http://web.worldbank.org>

World Trade Organisation. [www.wto.org/](http://www.wto.org/)

## Selected References

- Balassa, B. (1989) "Outward Orientation", in H. Chenery and T.N. Srinivasan (Editors), *Handbook of Development Economics*. Amsterdam: North-Holland, pp. 1645-89.
- Bhagwati, J.N. (1988) "Export-Promoting Trade Strategy: Issues and Evidence", *The World Bank Research Observer*, Volume 3, Number 1, pp. 27-57.
- Chang, H-J. (2005) *Kicking Away the Ladder: Development Strategy in Historical Perspective*. London: Anthem Press.
- Edwards, C. (1985) *The Fragmented World: Competing Perspectives on Trade, Money and Crisis*. London: Methuen.
- Emmanuel, A. (1972) *Unequal Exchange: A Study of the Imperialism of Trade*, London: NLB.
- Grabowski, R. (1994a) "Import Substitution, Export Promotion, and the State in Economic Development", *The Journal of Developing Areas*, Volume 28, Number 4, pp. 535-554.
- Haberler, G. (1987) "Liberal and Illiberal Development Policy", in G. M. Meier (Editor), *Pioneers in Development*. Second Series, Oxford: Oxford University Press.

- Hudson, M. (1992) *Trade, Development and Foreign Debt*. Volume I, London: Pluto.
- Jung, W.S. and P.J. Marshall. (1985) "Exports, Growth and Causality in Developing Countries", *Journal of Development Economics*, Number 18, pp. 1-12.
- Kozul-Wright, R. (1995) "Transnational Corporations and the Nation State", in J. Michie and J.G. Smith (Editors), *Managing the Global Economy*. Oxford: Oxford University Press.
- Krueger, A.O. (1985) "The Experiences and Lessons of Asia's Super Exporters", in A. O. Krueger, and V. Corbo (Editors), *Export Oriented Development Strategies: the Success of Five Newly Industrialising Countries*. Boulder, Colorado: Westview Press.
- Krugman, P. (1984) "Import Protection as Export Promotion: International Competition in the Presence of Oligopoly and Economics of Scale", in H. Kierzkowski (Editor), *Monopolistic Competition and International Trade*. Oxford: Clarendon.
- Krugman, P. (1998) "Ricardo's Difficult Idea: Why Intellectuals Don't Understand Comparative Advantage", in G. Cook (Editor), *The Economics and Politics of International Trade*. London: Routledge.
- Lal, D. and S. Rajapatirana. (1987) "Foreign Trade Regimes and Economic Growth in Developing Countries", *World Bank Research Observer*, Volume 2, Number 2, pp. 189-217.
- Leontief, W.W. (1954) "Domestic Products and Foreign Trade: The American Capital Position Re-examined", *Economia Internazionale*, Volume 7, pp. 3-32.
- Liang, N. (1992) "Beyond Import Substitution and Export Promotion: A New Typology of Trade Strategies", *The Journal of Development Studies*, Volume 28, Number 3, pp. 447-472.
- Marx, K. (1867) *Capital*, Volume I. London: Penguin, 1999.
- Michaely, M. (1977) "Exports and Growth: An Empirical Investigation", *Journal of Development Economics*, Volume 4, Number 1, pp. 49-54.
- Pack, H. (1992) "Learning and Productivity Changes in Developing Countries", in G.K. Helleiner (Editor), *Trade Policy, Industrialisation, and Development*. Oxford: Clarendon.
- Pritchett, L. (1996) "Measuring Outward Orientation in LDCs: Can it be Done?", *Journal of Development Economics*, Volume 49, Number 2, pp. 307-335.
- Rodriguez, F. and D. Rodrik. (2000) "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross National Evidence", in B. Bernanke and K. Rogoff (Editors), *Macroeconomics Annual 2000*. Cambridge, Mass: MIT Press for NBER.
- Santos-Paulino, A.U. (2005) "Trade Liberalisation and Economic Performance: Theory and Evidence for Developing Countries", *The World Economy*, Volume 28, Number 6, pp. 783-821.
- Schydrowsky, D.M. (1984) "A Policymaker's Guide to Comparative Advantage", *World Development*, Volume 12, Number 4, pp. 439-449.
- Shaikh, A. (1999) "Real Exchange rates and the International Mobility of Capital", *Working Paper Number 265*, New School University.
- Singer, H.W. (1984) "The Terms of Trade Controversy and the Evolution of Soft Financing: Early Years in the UN", in



Turan Subasat  
Department of Economics & International  
Development, University of Bath,  
UK  
t.subasat@bath.ac.uk

- G.M. Meier and D. Seers (Editors), *Pioneers in Development*. New York: Oxford University Press for the World Bank.
- Streeten, P. (1982) "A Cool Look at "Outward-Looking" Strategies for Development", *The World Economy*, Volume 15, Number 2, pp. 159-169.
- Subasat, T. (forthcoming) "Do Liberal Trade Policies Promote Trade Openness?", *International Review of Applied Economics*.
- Subasat, T. (2002) "A Political Economy Critique of the Ricardian Comparative Advantage Theory", *Research in Political Economy*, Volume 20.
- Subasat, T. (2003) "What Does the Heckscher-Ohlin Model Contribute to International Trade Theory?: A Critical Assessment", *Review of Radical Political Economics*, Volume 35, Number 2.
- Stolper, W.F. and P.A. Samuelson (1941) "Protection and Real Wages", *Review of Economic Studies*, Volume 9, pp. 58-73.
- Trefler, D. (1995) "The Case of the Missing Trade and Other Mysteries", *American Economic Review*, Volume 85, Number 5, pp. 1029-1046.
- Wood, A. (1994) *North-South Trade, Employment and Inequality: Changing Fortunes in a Skill-Driven World*. Oxford: Clarendon Press.
- Yaghmaian, B. (1994) "An Empirical Investigation of Exports, Development, and Growth in Developing Countries: Challenging the Neoclassical Theory of Export-Led Growth", *World Development*, Volume 22, Number 12, pp. 1977-1995.

## Future of Work

*John Burgess*

### **Introduction**

This discussion of the future of work will look at five key issues. The first is what do we mean by work? The second is whether it is possible to sustain sufficient work for all in the future. Third, there is the key issue of whether work should be available for all, especially those who do not wish to perform work. Fourth, there are the conditions and rewards associated with work. Finally, can the systems of regulation and representation associated with work be sustained or will they be replaced by new forms of regulation and voice? The discussion will consider work issues in the context of the advanced Western (OECD) economies. This is not to dismiss the significance of work elsewhere but it reflects the mainstream analysis of the models, forms and regulation of work that prevails across many of the OECD economies.

### **The Construction of Work**

Work is socially constructed. It has different meanings in different communities and in different locations. Likewise the purpose of work spans from punishment and necessity through to providing religious salvation. If we are to consider the future of work the starting points are the meaning and purpose of work.

Defining work is a difficult task. Essentially work is constructed in a social context. What we mean by work depends on the typology or reference points we construct for identifying work. In OECD economies work is officially defined in terms of the market exchange of labour

services. Defining and measuring work is located within the labour force framework that accords priority to employment over unemployment and to labour force participation over labour force absence (Australian Bureau of Statistics 1995). In this context one hour of employment per week is sufficient to classify a person as being employed and thus being in the labour force. The activity that is classified as employment involves the exchange of labour services for payment or in kind, or in a family business.

The application labour force framework results in work being defined in market terms and as such it excludes non market labour input. The activities of volunteers and within household labour are excluded by the market orientation of the official typology. Within the market value of work is measured in exchange terms and participation in work is given precedence over non participation in work for the purposes of population classification of those with work and those without work. Those engaged in schooling and education are not at work, as are those who engage in voluntary community activities, and those who engage in involuntary and unpaid activities through programs of conscription and incarceration. We can thus talk about work in strictly market terms or work as a “broader social activity where relationships do not lend themselves easily to economic measurement even though they require skill and the expenditure of time” (le Guidec 1996:645).

Work has not always been constructed and defined in such terms. In many communities around the globe work has a completely different meaning and construction from that found in many Western economies. The division between remunerated and unremunerated work and

between market and domestic work is not distinct where the family is the major form of organisation for purposes of production (Moore 1995). Also work as we currently know it was not always organised in its current form. Meda (1996) reminds us that work has passed through many stages of different meanings and constructions through history. The market classification of work also contains a gender bias against domestic work and “the labour of social reproduction” (Moore 1995:657).

The social construction of work in OECD economies will be challenged. The present construction has a strong gender and a market bias. Recognising arrangements outside of the traditional labour force framework as constituting work and integrating these arrangements into reward and welfare systems will be major challenge of the future.

### **Work for All in the Future?**

The future of work across Western economies has been subject to extensive review over the past 30 years. There is the debate over whether there is sufficient work for all since the end of the “Golden Age” of capitalist expansion (Lipietz 1992). Mass and sustained unemployment has become a permanent feature of most Western economies over this period. Unemployment remains entrenched despite improved growth in some economies throughout the 1990s. Most economies have moved into a post industrial stage since virtually all new jobs being generated are located in the service sector. This has led analysts to consider the regulation and production context in which jobs are being generated and whether a post industrial paradigm has replaced an earlier system of mass consumption and mass production (Fordism) supported by extensive systems

of government support for full employment (Lipietz 1999).

The permanence of unemployment and its uneven demographic distribution remains a challenging issue for all OECD governments. There is pessimistic view associated with writers such as Rifkin (1995) who argue that the growing capital intensity of work together with the globalisation of production are job displacing. Technology and machines can perform many of the functions previously carried out by paid labour. The nature of the production systems will not require as much labour input as previously. Production can easily be relocated to locations that provide lower labour costs. In this context the future of work for many will be in terms of providing voluntary labour.

There are also those who see an ongoing struggle to reduce unemployment and a growing segmentation of the workforce between the highly skilled and the unskilled. There will be work but there will be growing polarity of rewards and choices (Reich 1991). The workforce will become polarised between the high paid and the mobile and the low paid personal service workers who support the well off. In both cases there is a proclamation of the inability of the economic system to sustain work but also a prediction of growing inequality in terms of choices, rewards and conditions associated with work.

Location is increasingly transient in terms of job security and jobs in Western economies are found in the services. Mass consumption still supports the production system but jobs are now removed from industrial production and located across a range of diverse service industries ranging from personal and caring services through to business and computing services. The

nature of work in terms of the occupational profile of the workforce and the content of jobs is changing. However, are systems of work organisation and control over work fundamentally different from those associated with mass production?

Here we have two opposing views. First, neo Fordist views that argue that systems of control and co-ordination are fundamentally the same despite shifts in the industrial and occupational distribution of the workforce. One of the famous manifestations of this is the description of call centre work as “the assembly line in the head” (Taylor and Bain 1999). Second, those views that suggest we are witnessing a new production paradigm with an emphasis on teams as opposed to production lines, product quality as opposed to product homogeneity and separate production locations as opposed to a single production location. These are associated with arguments supporting flexible specialisation and post Fordist views of the production system (Mathews, 1989). Work is less secure, more mobile and linked to specialist skills. The nature of a career no longer single employer or single occupation focussed, it is related to following production and skill evolution across organisations, occupations and locations.

There is also an optimistic school who see future jobs for all, especially in services. The nature of jobs will change but there will be continued job availability, especially if labour markets are deregulated and impediments to labour matching are removed. This can be found in the OECD Jobs Study and by analysts associated with the OECD (Andrieu 1999; Goldfinger 1999). For many neo liberal governments there is sufficient work for all so long as there is both nominal and relative wage

flexibility. Unemployment is a manifestation of labour market imperfections, a high natural rate of unemployment view and there is no reason to believe that sufficient jobs for all would not be generated if markets were allowed to function effectively (OECD 1994).

### **What is the Purpose of Work?**

Work is connected to survival and reproduction. The gathering of food, the provision of shelter and clothing and the care of those unable to defend themselves are instinctive aspects of work. With the development of exchange through the market work itself becomes a commodity. The division of labour allows productivity to increase and for surpluses to be traded. Early forms of social organisation linked work to tithes and responsibilities to a system of hierarchy and obligation. Through markets work could be purchased and sold, however the provision of labour services also required the provider. That is, the provider and the service went together.

Increasingly work is not providing a link to minimum living standards. Across many OECD economies, even in the USA, the spectre of working poverty is emerging as a combination of very low pay, very short periods of employment engagements and short average working hours (Mishell and Schmitt 1995). The ability of work to provide an access mechanism to exercising effective market choice is being eroded. If work is no longer a mechanism for accessing minimum living conditions then it assumes other functions such as a disciplining mechanism for the unemployed through workfare type arrangements, for character building through conscription or as punishment through the prison system.

Work also gives identity and status. Categories of work may be accorded professional status and have conditions governing entry and the way work is performed. Work also has in the past been associated with secret societies and rituals (Sonnenscher 1989), and it has been a means for mobilising labour through trade unions and political movements linked to improving the conditions associated with work. Through work identity, class and consciousness can be established and work becomes the basis for social and political mobilisation (Hobsbawm 1959).

Work also has a spatial dimension. Work can be identified with location and ideas of community can be constructed around predominant employment arrangements (Dicken 1992). There have been the coal mining villages (South Wales, the Appalachians), shipbuilding towns (Middlesbrough and Bremen), financial cities (London and New York), cities based on public administration (Washington and Canberra), cities based on leisure (Monte Carlo and Las Vegas) and cities based on developments in the information and communications technology sectors (San Francisco). However, work, place and identity are also momentary and tenuous as many cities and regions have experienced the decline and disappearance of industries and jobs that established local identity (Dicken 1992).

Should the logic of work as providing meaning, identity and character be accepted? Not all forms of work are officially classified as work nor are they rewarded through the market. Tending a garden, caring for children or the elderly, organising community cultural and sporting activities, philosophising and debating, surfing, bikeriding and walking are ubiquitous but are not work, unless

someone pays you to carry out these activities. The link between work and reward, and the dichotomy between work and leisure will be increasingly questioned as populations age and service and caring work becomes the most predominant form of market work.

Work cannot be disconnected from the logic and sustainability of the underlying production systems. The ecological movement and the spectre of global warming have confronted the construction and logic of the Capitalist system of accumulation and its assumption of ongoing material progression. The use of non renewable resources, the depletion of eco systems and species and the congestion of cities does challenge the production system as a force for civilisation and progress (Hamilton 1994). If this is not sustainable then it means that jobs associated with it are also not sustainable, in turn this means that production has to be carried out with a different logic and in different ways. The nature of jobs and the underlying construction of a job moves away from a market and a material focus towards sustainability, non market and a community focus.

### **Work, Rewards and Welfare**

Closely associated with the above is the tentative and changing links between work and reward systems, and work and welfare. Working arrangements are becoming more fragmented and more heterogenous. The full time ongoing employment model linked to a few discrete employers is fast disappearing. It is being replaced by a diverse array of employment arrangements including part-time work, fixed term work, temporary work, homework and contract work. Single employers are being replaced by multiple employers, the single

workplace is being replaced by multiple workplaces and the home, employment continuity is being replaced by employment discontinuity and employment security is being replaced by employment insecurity (Standing 1999). As a result of this for many workers it is difficult to sustain predictable income flows linked to work and it is difficult to plan or to borrow funds without income predictability or security.

The model of social reproduction linked to the male breadwinner model of employment is being replaced by more diverse work and family arrangements (Crompton, Gallie and Purcell 1996). Linking minimum wages to family needs is more difficult to sustain as is linking welfare arrangements to a standard model of employment and household division of labour. Women's labour force participation rates continue to increase as has the average years of their participation in the labour force. Fertility rates are declining and multiple employment participating households co-exist with non employment participating households (Mishell and Schmitt 1996).

Working arrangements are becoming more diverse as are the forms of labour force participation within and between households. The challenge will be to establish a link between market employment and income that is predictable and secure and provides a basis for making such decisions as forming a relationship and acquiring assets. Following, the next challenge will be to establish a system of income support that recognises the emerging reality of diversity in employment arrangements. Rewards and the continuity of employment will not be sufficient for many to generate savings that provide a cushion for illness or for income

support in old age. The challenge for many governments will be to develop effective redistributive mechanisms in the face of the diversity and polarisation of employment arrangements and rewards associated with work.

Work remains the major source of personal income. Without work there is a need to access savings or to have access to forms of redistributive transfers. While political democracy is the prevailing ideology there is no imperative for economic democracy. Participation in the market requires effective demand, if work is unavailable or rationed then in the absence of redistributive mechanisms there can be no economic democracy.

### **The Regulation of Work**

The regulation of work is becoming more problematic for four reasons. First, mechanisms of voice and collective representation are declining (Standing, 1999). The trade union density has declined and it is questionable as to whether voice can be provided through non union forms of collective organisation. The rise of US human resources management and its supplanting of unions through rules and policies at the workplace, through bonus systems and share options means that it is increasingly difficult to mobilise and organise labour. Second, the public sector has been transformed through privatisation, corporatisation, contracting out and the adaptation of private sector employment practices (Supiot 1996). In many countries the public sector is no longer a model employer and public sector unions, for so long central to many national union organisations, face major representational problems as employees are transferred to the private sector or public sector jobs disappear. Third, the

fragmentation and diversity of work means that it is difficult to organise and mobilise many workers, and it is equally difficult to apply regulatory norms to their employment arrangements. This applies to sub-contractors, agency workers and casual workers. Importantly, many employment arrangements have an ambiguous status and hence the application of regulations is problematic. The temping industry is expanding, it offers employers cost and labour flexibility advantages, and in many countries the ambiguity of the employment relationship renders employment regulation difficult (Brewster et al 1997; Carre & Tilly 1998). Waged work is underpinned by a system of labour regulation that is having difficulty confronting the removal of collective norms and the fragmentation of employment arrangements (Castel 1996). Fourth, the development of globalised businesses and globalised business strategies means that production and jobs become highly mobile. The quest for cost savings, strategic location and favourable tax treatment results in national regulations being ineffective. The globalisation of the call centre industry means that companies are internally restructuring their organisations and spatially restructuring their organisations. Jobs can be removed from head offices and from the out of the country altogether (Dicken 1992). National systems of employment regulation are further undermined. In this context it is not surprising that workers across the OECD feel increasingly insecure about their jobs (Heery and Salmon 2000; Standing 1999).

Employment norms and the effective regulation of such norms is becoming increasingly problematic. The EU has been grappling with systems of super national employment codes and systems of labour

regulation that recognise the diversity and fragmentation of employment arrangements (Supoit 2000). Likewise the ILO has to consider the application of labour standards in the context of very footloose industries and very capital friendly governments (Standing 1992). There are proposals for re-regulation of the employment relationship, for extending citizenship into the employment relationship and for reconstructing work that gives recognition to lifestyle diversity (Muckenberger 1996).

How far should the removal of regulatory norms and the de-collectivisation of the workforce be allowed to progress? There are concerns at two levels.

At the first level, the process facilitates growing inequality within the workforce and breaks the nexus between work and basic living conditions. If work cannot be made to pay then a policy emphasis on supply side responses becomes meaningless. It also implies that the State will have to provide transfers to support work and to implicitly subsidise those businesses with low paid workers.

At the second level, in an ageing society context there is an imperative from governments to get citizens to fund their own retirement through voluntary and compulsory contributions from their wages towards a retirement fund (OECD 1998). However, where work is fractional, irregular and low-paying it becomes impossible to sustain regular contributions to a retirement fund that provide for the most rudimentary retirement income. High unemployment rates, non-standard employment arrangements and a large number of low paying jobs are not conducive to a broad based self-funded retirement program.

## Conclusions

The nature and forms of work are changing, as is the social construction of work together with the meaning of a career. The production and reproduction systems behind work are being challenged and the ability to sustain market work for all is questionable. The links between work and rewards and work and welfare systems are becoming more tenuous. The ability to regulate work and mobile workers is becoming more problematic. Work will be sustained, but in different forms and with different foundations. Should work be available for all? What should be defined as work? What about those who do not have officially sanctioned work, what is their place in the community? These issues will persist and pose significant challenges for communities and policy makers.

## Selected References

- Andrieu, M. (1999) "A Better Future for Work?", *OECD Observer*, Summer, 53-54.
- Australian Bureau of Statistics (1995) *Measuring Employment and Unemployment*. Canberra: ABS, Catalogue 6279.0
- Brewster, C., L. Mayne and O. Tregaskis. (1997) "Flexible Working in Europe", *Journal of World Business*, Volume 32, Number 2, pp. 133-151.
- Carre, F. and C. Tilly. (1998) Part-time and Temporary Work: Flexibility for Whom? *Dollars and Sense*, January-February, Number 215, pp. 22-25.
- Castel, R. (1996) Work and Usefulness to the World, *International Labour Review*, Volume 135, Number 6, pp. 615-622.
- Crompton, R., D. Gallie and K. Purcell. (1996) *Changing Forms of Employment*. Routledge, London.
- Dicken, P. (1992) *Global Shift: the Internationalisation of Economic Activity*. London: Paul Chapman.
- Goldfinger, C. (1998) *The Future of Work*. UNESCO Courier. December, 34-36.
- Hamilton, C. (1994) *The Mystic Economist*. Canberra: Willow Park.
- Heery, E. and J. Salmon. (2000) *The Insecure Workforce*. London & NY: Routledge.
- Hobsbawm, E. (1959) "The Labour Sects", in E. Hobsbawm (Editor), *Primitive Rebels*. Manchester: Manchester University Press, pp. 126-149.
- Joyce, P. (1989) "The Historical Meanings of Work: An Introduction", in P. Joyce (Editor), *The Historical Meanings of Work*. Cambridge: Cambridge University Press, pp.1-30.
- Le Guidec, R. (1996) "Decline and Resurgence of Unremunerated Work", *International Labour Review*, Volume 135, Number 6, pp. 615-622.
- Lipietz, A. (1992) *Towards a New Economic Order*. London: Polity Press.
- Lipietz, A. (1997) "Economic Restructuring: the New Global Hierarchy", in P. James, W. Veit and S. Wright (Editors), *Work of the Future*, Global Perspectives. Sydney: Allen and Unwin, pp. 45-65.
- Lowe, G. (1998) "The Future of Work", *Industrial Relations*, Volume 53, Number 2, pp 235-257.
- Mathews, J. (1989) "From Post Industrialism to Post Fordism", *Meanjin*, Volume 38, Number 1, pp. 139-152.
- Meda, D. (1996) "New Perspectives on Work as Value", *International Labour Review*, Volume 135, Number 6, pp. 633-643.
- Mishel, L. and Schmitt, J. (1996) *Beware the US Model*. Washington DC: Economic Policy Institute.



- Moore, H. (1995) "The Future of Work", *British Journal of Industrial Relations*, Volume 33, Number 4, pp. 657-678.
- Muckenberger, U. (1996) "Towards a New Definition of the Employment Relationship", *International Labour Review*, Volume 136, Number 6, pp. 683-695.
- OECD (1994) *The Jobs Study*. Paris: OECD.
- Reich, R. (1991) *The Work of Nations*. New York: Simon and Schuster.
- Rifkin, J. (1995) *The End of Paid Work*. New York: Putmans.
- Sonnenscher, M. (1989) "Mythical Work: Workshop Production and the Compagnonnages of Eighteenth Century France", in P. Joyce (Editor), *The Historical Meanings of Work*. Cambridge: Cambridge University Press, pp. 64-98.
- Standing, G. (1999) *Global Labour Flexibility*. London: Macmillan.
- Supiot, A. (1996) "Perspectives on Work: Introduction", *International Labour Review*, Volume 135, Number 6, pp. 603-614.
- Taylor, P. and P. Bain. (1999) "An Assembly Line in the Head: Work and Employment Relations in the Call Centre", *Industrial Relations Journal*, Volume 30, Number 2, pp. 101-117.

*John Burgess*  
*Newcastle Graduate School of Business*  
*University of Newcastle*  
*Australia*  
*John.Burgess@newcastle.edu.au*

## Game Theory

Roger Ashton McCain

### **Introduction**

What can the theory of games have to do with governance and public policy? As 2005 Nobel Laureate Robert Aumann (2003) points out, the term “game theory” is a somewhat misleading term for what might be more accurately called “a theory of interactive decisions.” Since interactive decisions largely make the world of politics and social action what it is, and public policy and governance are activities designed to influence decisions so that the ultimate result of their interaction is better on the whole, we may suppose that game theory is indeed fundamental for governance and public policy.

### **Cooperative and Noncooperative Games**

Game theory was largely founded by von Neumann and Morgenstern (1944) with their book *The Theory of Games and Economic Behavior*. In this book they presented a general solution to the simplest case of interactive decisions, “two person constant sum games,” that is, interactions in which any one decision making agent can increase his own payoff only by causing a one-for-one decrease in the payoff of the other. However, they did not attempt to extend that solution to games of more than two persons or non-constant sum (“win-win” and “lose-lose”) games, since they felt that in these more general games the formation of groups (“coalitions” in game theory terms) would be essential. In “win-win” games, for example, rational agents would form coalitions to realize the win-win potential. In games of three or more persons, some would form coalitions to exploit others. However, John Nash (1950a, 1951) proved that the von

Neumann and Morgenstern solution for two-person constant sum games could be extended to n-person and nonconstant sum games (for which he was honored with the 1994 Nobel Memorial Prize in Economics) and proposed, in the “Nash Program,” (Nash 1950b, 1953; Serrano 2003) that coalitions should be considered as equilibria of some noncooperative bargaining game. Out of this came the distinction of cooperative and noncooperative solutions to games, which is of central importance for governance and public policy.

As an instance, consider the following Public Goods Contribution Game. Two agents, A and B, are each initially endowed with 5 units of private wealth. Each has the option of contributing three units of private wealth to the creation of a nondepleteable, nonexclusive *public good*. For each 3 units of private wealth contributed, 2 units of the public asset are created. The *payoff* to an agent is the sum of the public good and the private wealth he has remaining at the end of the game. Thus, for example, if both contribute their three units, each ends the game with 2 units of private wealth, but four units of the public good are created, so the payoff to each agent is  $2+4=6$ .

For each agent there are two strategies: contribute or don't contribute. Table 1 shows the strategies chosen by A and B and the resulting payoffs. The table is read as follows. Each row corresponds to one of agent A's strategy choices, and each column corresponds to one of agent B's strategy choices. Thus, for example, considering “agent A contributes but agent B does not” as a candidate solution, we find the results of those strategy choices in the upper left corner where the first row and the second column intersect. The results are the payoffs to the two agents. Following a common convention, the first number is the payoff to

agent A and the second number is the payoff to agent B. Table 1 is an example of a *game in normal form*, an innovation of von Neumann and Morgenstern, who showed that (with sufficiently complicated notation) any game can be represented in normal form.

*Table 1.* The Two Person Public Goods Contribution Game in Normal Form

		B	
		contribute	don't
A	contribute	6,6	4,7
	don't	7,4	5,5

We first consider the *noncooperative* outcome of a game such as this. “Noncooperative” means that there is no coordination or correlation between the agents’ decisions: each agent assumes that the other agent’s decision is independent of his own. Suppose we have a list of strategies, comprising one strategy for each player. This list of strategies is a *Nash equilibrium* if no agent can benefit by unilaterally switching to another strategy, while the other agent or agents do not.

Each agent makes a conjecture as to which strategy the other may choose or how the other agent may choose his strategy; but does so assuming that the others’ choice is independent of his own and that the other agent makes the reciprocal assumption. In this case it is not hard for agent A to guess what B will do. A can observe that in case A himself chooses “contribute,” then B will have 7 rather than 6 by choosing “don’t;” while if A chooses “don’t contribute,” B has 5 rather than 4 by choosing “don’t.” Therefore, “contribute” is a *dominated strategy* for B, in the terms of game theory, and A can eliminate the dominated strategy from consideration, assuming that B, a rational agent, will never choose a

dominated strategy. But B can reason the same way with respect to A and can eliminate the dominated strategy for A, as well. The lower right cell, corresponding to a case in which neither person contributes and none of the public good is produced, remains as the only outcome that both agents feel is possible given the other person’s rationality. It is easy to check that this outcome is a Nash equilibrium. If either agent switches to “contribute,” his payoffs drop from 5 to 4. This the noncooperative solution to this game.

We now consider the *cooperative solution* for the game. “Cooperative” means that the two agents are able to form a coalition and choose strategies as a group, realizing any opportunity to benefit by coordinating or correlating their strategies. Such an agreement is enforceable in some way or constitutes a binding commitment on all members of the coalition. The group can choose among four strategy possibilities – “both contribute,” “A contributes but B does not,” “B contributes but A does not,” and “neither contributes.” The two unsymmetrical possibilities—“A contributes but B does not,” and “B contributes but A does not”—can be ruled out immediately. Rather than agree to “A contributes but B does not,” A can do better by dropping out of the group and going it alone by choosing “don’t contribute.” Similarly, B’s interests rule out “B contributes but A does not.” As between the symmetrical possibilities, “both contribute” and “neither contributes,” both are better off when both contribute. If one tries to drop out of the coalition (to *defect* in the usual terminology of game theory) and not contribute for a payoff of 7, the coalition breaks down and the other refuses to contribute as well, leaving the defector with a payoff of 5 instead. Thus “both contribute” is the cooperative solution of the game.

We note that in the public goods game, the cooperative and noncooperative solutions disagree, so that both agents are worse off in the noncooperative solution. Since moreover the noncooperative solution is a dominant strategy solution, the public goods game is a “social dilemma.” (Dawes 1980) The famous Prisoner’s Dilemma (Poundstone 1992; Rapoport and Chammah 1965) is the best known and founding case of a social dilemma. Social dilemmas provide a category of games in which a mutually desirable outcome is opposed by very strong individualistic motives and so cannot be brought about without some enforcement of the common strategy. As such it is widely considered a paradigmatic case for positive public policy. (e.g. Barash 2004) In many applications, the reasonable objective of governance and public policy is to implement cooperative solutions in cases when pure voluntarism would lead to less-efficient noncooperative solutions.

This conception can be criticized as incomplete. When we argue that some social problem arises from a social dilemma, so that government intervention is justified, we take government (or some governance institution capable of imposing a binding agreement) as a *Deus ex Machina*. The analysis seems to have no application to governance institutions whose powers are not sufficient to impose a binding agreement or cooperative solution. On the other side, it considers only a narrow range of noncooperative games and Nash equilibria. Coalitions may not always be excluded in noncooperative games. To better illustrate the range of noncooperative game analyses, we now digress to explore a game-theoretic model of terrorist strategies aimed at destabilization of social relations that might otherwise be stable.

## **Noncooperative Games: The Logic of Destabilization**

Since the 1950’s, game theory has had two broad and largely independent branches: the theory of noncooperative games (of which Nash’s equilibrium concept is an example and certainly among the most influential of concepts) and the theory of cooperative games. The theory of noncooperative games has had many successes and its influence has tended to eclipse that of the theory of cooperative games, as witness the awards of the 1994 and 2005 Nobel Memorial Prizes in Economics for contributions to noncooperative game theory. The 1994 prize to Harsanyi, Nash and Selten honored “their pioneering analysis of equilibria in the theory of non-cooperative games” while the 2005 prize honored Schelling and Aumann for helping to answer “Why ... some groups ... succeed in promoting cooperation while others suffer from conflict?” (Bank of Sweden 1994, 2005).

This aspect of noncooperative game theory can be illustrated by an extensive discussion of terrorism. Probably no issue in public policy is more crucial and troubling than terrorism. However, terrorism may seem to be an implausible application for a theory that, like game theory, assumes rationality. We are tempted to the view that “terrorists are just irrational” or “they just hate our freedom.” But, however emotionally satisfying that conception may be, and even if it is in some sense true, it does not help us to anticipate, respond to, nor to prepare for future attacks. However “insane” the objectives of terrorists may be, they seem to choose their strategies to promote those objectives with a great deal of instrumental rationality, and that in turn means that their tactical choices can be analyzed in game theoretic terms. In particular, noncooperative game theory

seems appropriate, since the motives of the terrorists and their victims are wholly opposed. Moreover, we will argue that many instances of terrorism have a common logic, in that the terrorists aim at destabilizing what might otherwise be a stable noncooperative equilibrium among the major actors. This will be illustrated by two historical examples as well as the current terrorist challenge from al-Qaeda and similar groups.

Terrorists rarely “win,” but there are a few exceptions. Consider Gavrilo Princip, a supporter of Yugoslav (Greater Serbian) nationalism in 1914. It was he who assassinated the Archduke of Austria, setting off the First World War. Among the outcomes of that war was the dissolution of the Austro-Hungarian Empire and the creation of Yugoslavia by the addition of Slovenia and Croatia (formerly Austro-Hungarian provinces) and Bosnia-Herzegovina (formerly occupied by Austria-Hungary) to the Yugoslav monarchy. Gavrilo Princip was a terrorist who won.

Prior to World War I, the great powers of Europe were caught up in a network of hostile relationships and treaties, both secret and public, that had remained more or less stable for some time. However, the dissolution of the Ottoman Empire in Europe had given rise to some new states on the southern border of Europe and their relationships among themselves and to the great powers remained undefined, giving rise to the Balkan Wars of 1912-13. (Burn 1989:1057) Serbia, in particular, was an Orthodox Christian Slavic nation with a close relationship with Russia; and Russia was committed to war, if necessary, to protect Serbia. But Serbia was particularly hostile to Austria-Hungary, which had annexed Bosnia-Herzegovina, a former Ottoman territory with a substantial

population of Serbs, and had long incorporated the Slavic (but mostly Catholic) peoples of Croatia and Slovenia. While Russia would not tolerate an Austro-Hungarian attack on Serbia, Austria-Hungary could not improve its relations with Serbia without ceding territory; but such cessions might prove (as in fact they did) only the first step in the dissolution of the Austro-Hungarian state. Gavrilo Princip represented an extremist faction in Serbia that aimed to create a Greater Serbia through the incorporation of Slavic regions in the south of the Austro-Hungarian territory. (Burn 1989:1058)

The simplest game theoretic model of this incident represents it as a game in normal, that is, tabular form. At first suppose the only players in the game were Austria and Russia. We can analyze their interaction using the game in normal form, as with the public goods contribution game. The game is shown in Table 2. This table is read just as Table 1. The idea behind the table is that peace is best for both countries, but if there is to be war, the country that chooses peace (i.e. does not mobilize except in defense) is even worse off than if they had chosen war. In other words, if there is war, the aggressor is at an advantage. The numbers are arbitrary and quite hypothetical, but are chosen to express that idea and one other: Russia had little to lose by choosing war, but Austria had a great deal to lose.

*Table 2. Two-Person Game—Austria and Russia*

		Austria	
		war	peace
Russia	war	-5,-5	1,-7
	peace	-7,-3	2,2

As the table shows, this game is not a social dilemma. There are no dominated

strategies in the game. However, the game has a Nash equilibrium, and in fact it has two. For example, starting from the lower right “peace, peace,” suppose that Russia shifts to war (the row above). Then her payoffs are reduced from 2 to 1. Similarly, if Austria switches but Russia does not (to the last row in the left column) Austria’s payoff is reduced from 2 to -3. Thus, the lower right cell is a Nash equilibrium. But the upper left is also a Nash equilibrium, since if either power shifts from “war” to “peace” unilaterally, that power’s payoff is cut from -5 to -7.

While not based on any specific work of Schelling’s, this example is very consistent with the ideas for which he was honored in the 2005 Nobel Memorial prize in economics. If one of the two Nash equilibria attracts attention as being more likely than the other one, then both will choose that equilibrium and their expectations prove to be mutually self-confirming. (Such a Nash equilibrium is often called a *Schelling point*.) In this case, the lower right equilibrium attracts attention for two reasons: first, it is better for both countries, and second, it continues the status quo of the previous few years. Peace may well be a Schelling point.

In these circumstances, an existing peace should be stable, but this is not the whole story, because there were other players in the game. We will use an extension of the tabular normal form representation to represent the interaction of both countries with Serbia as a noncooperative three-person game. Table 3 shows a three-person game with Serbia as the third player. To read this table is just a little more complicated than reading Tables 1 and 2. First, the two panels of the table correspond to the strategy alternatives of the third party: in this case the Serbian extremists. In this case, the left

panel is applicable if the Serbian extremists choose no terror, while the right panel is applicable if Serbia chooses terror. Within each panel, the table is read as before, with each row corresponding to a strategy for Russia and each column to a strategy for Austria-Hungary. As before, each cell is a candidate solution, showing the payoffs to Russia, Austria-Hungary, and Serbia in that order. (This is not to imply that the decisions are taken in a particular sequence, and we will treat them as simultaneous).

Table 3. Three-Person Game with Serbia

		Serbia			
		No Terror		Terror	
		Austria		Austria	
		war	peace	war	peace
Russia	war	-5,-5,5	1,-7,5	-5,-5,4	1,-7,4
	peace	-7,-3,5	2,2,0	-7,-3,4	2,-4,-1

The payoffs to Russia and Austria in the left panel are as before, but in the right panel, Serbian terrorism creates an intolerable situation for Austria, even in peacetime. The payoffs for Serbia reflect the assumption that Serbia can gain only in case one of the other powers chooses war, but also that terrorism costs the terrorist party something. This three-person game has three Nash equilibria, shown by the shaded cells, but not all Nash equilibria are equal. We may be able to *refine* the Nash equilibrium criteria to set off one of the equilibria as more probable. Such *refinements* play a large role in the literature of non-cooperative game theory. (Harsanyi & Selten 1987.)

Suppose that Serbia chooses “no terror.” This leaves Austria and Russia back in the original two-person game, Table 2, in which, as we have seen, peace may be stable; but for Serbia the payoff in the case of peace is zero. War is also a possibility in that game,

so Serbia has some hope of its payoff of 5, but no certainty of it. If Serbia chooses terror, Austria and Russia are placed in the two-person game shown in Table 4. This game has only one Nash equilibrium, the upper left cell in which Austria and Russia both choose war. Thus, Serbia can eliminate the uncertainty and assure herself of her objective, the payoff of 4 in the table, by a strategy of terrorism. It is correct to say that Serbian terrorism destabilizes the peace, and it is also correct to say that Serbian terrorism succeeded by provoking Austria to a war that resulted in Austria's destruction.

*Table 4. Two-Person Game if Serbia Chooses Terror*

		Serbia—Terror	
		Austria	
		war	peace
Russia	war	-5,-5,4	1,-7,4
	peace	-7,-3,4	2,-4,-1

The three-person game in Table 3 has three Nash equilibria, the three shaded cells, but only one has the further property of being a “perfect equilibrium,” that is, failsafe for Serbia, and it is the war, war, terror equilibrium shaded a darker gray. This equilibrium describes what did in fact occur. The key lesson is that, in this instance, the terrorists won by precipitating a larger conflict that achieved the objective they themselves did not have the force to attain. This is the logic of destabilization.

Let us consider one more example: the leftist terrorists of the 1960's-1970's, such as the Red Army Faction. Here, of course, we have a terrorist group that did not win. Our working hypothesis is that the terrorists' objective was to destabilize (as Prinzip did) what might otherwise be a stable equilibrium against their objectives. In Table 5 we see a

three-person game table with the Red Army Faction in the place of Serbia, the Social Democrats in the place of Russia, and the conservatives in the place of Austria. The payoffs are repeated from Table 3.

*Table 5. Red Army Faction*

		Red Army Faction			
		No Terror		Terror	
		Conservatives		Conservatives	
		civil war	no war	civil war	no war
Social Democrats	civil war	-5,-5,5	1,-7,5	-5,-5,4	1,-7,4
	no war	-7,-3,5	2,2,0	-7,-3,4	2,-4,-1

From its own point of view, the Red Army Faction faced a static situation (the Nash equilibrium at no war, no war, no terror) in which the moderates of the working class had a nonrevolutionary accommodation with the employing class. The objective of the revolutionary left extremists was to provoke the conservatives into retaliations that would have forced the working class to fight in self-defense. Their hope was then to lead a victorious revolutionary war and make themselves dominant in the state. The payoffs to civil war are probably exaggerated in this case—it may be that neither major player really gained anything from choosing civil war in any circumstances—but the purpose of the table is to reflect the situation as the Red Army Faction saw it, not necessarily as it was.

Taking the payoff table as assumed, however, the different outcome (by comparison with the previous example) could reflect the fact that the Conservatives and the Social Democrats formed a coalition. (The reference here is not to their

Grand Coalition in the Bundestag but to a less formal, more enduring joint commitment on their response to terrorism). Effective coalitions can form in noncooperative games when there are more than one Nash equilibria. (Bernheim et al 1987.) Effectively, the terms of the coalition were that terrorism should never be treated as a military challenge but always as a matter of criminal enforcement. Thus, the coalition chose the no war, no war strategy combination, making terrorism a losing strategy for the extreme left. Because no war, no war was a Nash equilibrium, there would be no need to enforce it, so the coalition policies could be realized even in a noncooperative interaction. In fact, Europe defeated the Red Army Faction by resisting the terrorists' claim to be something other than criminal defendants. By persisting in treating terrorism as a matter of criminal law enforcement, the European governments directly opposed the proximate objectives of the terrorists, and did so successfully. It may be that this behavior was sometimes cooperative rather than noncooperative – that in some periods a military response to terrorism would in fact have made one side or other better off, but that they remained committed to not supporting war measures despite their self-interest. It may be that the Red Army Faction miscalculated and the payoffs were not as shown here. What is clear is that, in the face of the persistent commitment of the main power centers in Europe to criminal enforcement rather than war measures against terrorism, the extremists were unable to mount a terror campaign sufficient to destabilize Europe.

The same logic could be applied to Russia, Austria and Serbia. A Russian-Austrian coalition on the terms that terrorist action could not be a *casus belli* might have prevented the First World War, but such a

coalition seems to have been unthinkable to them. The idea that a peace coalition without enforcement powers might nevertheless stabilize peaceful relations among the great powers underlay the League of Nations (see e.g. Macmillan 2002:ch.7) and undoubtedly also is a key idea for the United Nations. Unfortunately, this is effective only when peace is a Nash equilibrium, and peace may not always be a Nash equilibrium. When a great power chooses to destabilize the existing order, as in the 1930's, then the multiple-equilibrium analysis may not be applicable. If peace is not a Nash equilibrium then, on this view, a peace coalition without enforcement (i.e. federal government) powers will not be effective.

The logic of destabilization is also applicable to the terrorism of al Qaeda and similar groups. Perhaps al Qaeda perceives the world as shown in Table 6. The table is read as tables 4 and 5 and again repeats the same payoffs. Once again, the terrorists' objective would be to provoke the USA and its allies to a retaliation that would in turn force the Muslim moderates to fight (and unite) in self-defense.

Table 6. Al Qaeda

		Al Qaeda			
		no terror		terror	
		Muslim Moderates		Muslim Moderates	
		civil war	no war	civil war	no war
USA and allies	war response	-5,-5,5	1,-7,5	-5,-5,4	1,-7,4
	no war	-7,-3,5	2,2,0	-7,-3,4	2,-4,-1

The logic of destabilization suggests that the United States was mistaken in stressing war measures rather than (to the greatest extent possible) law enforcement measures in response to the terrorist attacks of 2001. It



also suggests that an effective coalition in this case necessarily would include the USA and the Muslim moderates, difficult as this might be for both sides. Of course, all of this assumes that the logic of destabilization applies to the current terrorist challenge. As we have seen, this approach is special, applicable only when the payoffs are enough like those in Table 6 so that there are multiple equilibria, including a peace equilibrium. Even if Table 6 captures the view of the world that terrorists hold, and that motivates them to choose terrorism, it may not be an accurate description of the actual world and almost certainly does not approximate the view of the world held by United States governments in the twenty-first century.

We have seen in this section that noncooperative equilibrium theory can 1) enable us to understand how even such seemingly irrational tactics as terrorism may be rationally chosen, and 2) illustrate how coalitions among sovereign players may affect outcomes even without means of enforcement of the terms of agreement, and the limitations of such coalitions. Nevertheless, it may be that noncooperative approaches do not capture the whole world of interactive decision theory.

### **Issues for the Noncooperative Approach**

The predominance of noncooperative game theory in public policy applications probably reflects two things. First, noncooperative solutions are observed in field observation and experiment often enough that they may seem, as a whole, better predictors of human behavior than cooperative solutions (Mailath 1998). Second, in part this influence reflects the pragmatic usefulness of noncooperative game theory for public policy, in posing the key problem that policy is needed to solve, as the Public Goods Contribution Game

illustrates. The game can readily be extended to a more realistic, large number of participants with similar results: that contributions are lower and all are worse off in the noncooperative than in the cooperative outcome. Such game examples show how voluntarism may fail to supply public goods, so that compulsory contributions via taxation and a public authority may improve the welfare of all. In many applications, the reasonable objective of governance and public policy is to implement cooperative solutions in cases when pure voluntarism would lead to less-efficient noncooperative solutions.

Industrial organization and antitrust economics are largely organized around a major exception to this. Economists think of price competition as noncooperative pricing behavior. If the firms in a market form a cartel, they are cooperating to more effectively exploit their customers, so in this case the objective of public policy would be to prevent the cooperative solution and to encourage the noncooperative one. However, when the customers are modeled as players in the game, the core, a cooperative solution concept, corresponds to competitive pricing; so this is really no exception at all. (Shubik 1959)

There are, however, difficulties to match each of these positives. Noncooperative equilibria are not uniformly empirical predictors of human behavior in game situations. The pragmatic usefulness of noncooperative equilibrium theory is bought at the cost of treating government as a *Deus ex Machina*, not a player in the game. This difficulty is all the more troubling if we attempt to extend game theory to analyze governance in a pluralized political situation, in which various organizations may play interacting governance roles. Finally, in Maskin's (2004) words, we live our lives in

coalitions. As we have seen, coalitions can occur in a noncooperative game, but only under very restrictive circumstances, so that noncooperative game theory really tells us very little about the formation and functioning of coalitions, which in turn arguably is a central topic for economic and public policy theory.

### Cooperative Games

Accordingly, we might look instead at cooperative game theory in the hope of more reliable insights on governance and public policy. Cooperative game theory always presupposes coalitions, so that some authorities prefer the term “coalitional game theory” to “cooperative game theory” (Serrano 2003). For this perspective, a coalition might be just that—a group of parties in a parliament—or a club, a business firm, a nation state, an international treaty organization, a church, a labor union, or any other grouping that can choose a common strategy. The outcomes may include bribes, or *side payments* in the language of game theory, that make it worthwhile for the player to adopt a strategy that, in the absence of a side payment, would be disadvantageous to the player. It is hard to imagine any issue of public policy that does not fit within this framework.

However, there are several competing conceptions of cooperative solution, and cooperative game theory is richer in deep unsolved problems than in proven solutions. In general, we may say that a candidate cooperative solution to a game would describe 1) a set of coalitions, or *coalition structure*, that is likely to form and 2) an assignment of outcomes to the participants in the coalitions, where the outcomes may be money payoffs or more complex objects some of which an agent may prefer to others. These questions might be answered

without reference to the specific choices of strategies and interactions, looking only at the overall advantage that each agent can obtain by adhering to a particular coalition, and that is the usual approach in cooperative game theory. Games represented in this way – with the payoffs dependent only on the coalitions formed, assuming that the strategies chosen are the best available – are said to be in *coalition function form* or *partition function form*. Different approaches to solution may focus more on determining which coalitions will form (*the endogenous formation of economic coalitions*, Carraro) or may focus more on the division of the spoils, and may draw different conclusions. Reconciliation of these objectives and approaches remains a work in progress (Maskin 2004).

For the two-person public goods dilemma, for example, there are only two possible coalition structures: either both players go it alone and act as *singleton coalitions* or they together form the *grand coalition* of all players. The grand coalition can realize total payoffs of 12, while each agent can assure himself of a payoff of 5. Thus, any assignment of payoffs (after side payments) from just under seven to just over five, for A, and the rest for B, will make both players better off than they would be without the coalition. Which assignment in this range will the two players agree upon? Put this way, the question seems to be one of bargaining. This was Nash’s (1950b, 1953) approach in his contributions to cooperative game theory and the foundation of the Nash Program to integrate cooperative and noncooperative game theory (Serrano 2003). Shapley (1953) originated a quite different approach to dividing the spoils that is still being developed by game theorists (e.g. Maskin 2004). The theory of the *core* of a game (Shubik 1959, Telser 1996) is focused

more on determining which coalitions are likely to form, relying on the concept of *coalition dominance*, which will be illustrated below.

It will not be possible to survey all of these solution concepts, and even a full discussion of one or the other would unavoidably be quite technical. Instead, we will explore a single example, drawing on the ideas of Riker (1962) and the concept of dominant and dominated coalitions. The formation of a coalition in a parliamentary system provides a good instance of a cooperative game, since coalition agreements are enforceable. If they lapse a new election must be called. The game among the parties is also a relatively simple kind of game, a majority game—in which a coalition able to cast a majority of the votes “wins” and can enact its program.

We will illustrate this by the political maneuvering in Germany following the election of 2005. The 2005 elections in Germany did not give either of the customary coalition partnerships a majority, and even the plurality was a little unclear. The voting strengths of the parties represented in the German Bundestag after the election are shown in Table 7.

*Table 7. German Bundestag Seats, late 2005*

Total Seats	614	
CDU	180	29.3%
CSU	46	7.5%
CDU-CSU	226	36.8%
SPD	222	36.2%
FDP	61	9.9%
Greens	51	8.3%
Left	54	8.8%

The partnership of the CDU-CSU (Christian Democratic Union and Christian Social Union) had the plurality, although no party alone had more votes than the SPD

(Social Democratic Party). Nevertheless, neither the previous governing coalition of the SPD and the Greens nor the traditional alternative coalition of the CDU-CSU with the FDP (Free Democratic Party) had a majority. This led to a period of negotiation as each side tried to construct a coalition with more or other than the traditional coalition partners (and with nobody admitting the Left officially to government). The outcome was, as many predicted, the agreement of the CDU-CSU and the SPD to form a “grand coalition.”

Why did this come about, and why did it come about only with such difficulty? To address this question, we borrow a concept from the branch of cooperative game theory known as *the theory of the core*. The concept we will use is *coalition dominance*. The idea of a dominant coalition is quite different from the idea of a dominant strategy, although there is a mathematical connection between the two. To illustrate this concept, consider the list in Table 8 of possible coalitions in the new German Bundestag. Following Riker, we consider only minimal winning coalitions, i.e. coalitions that can attain at least 51% of the vote, and that include only enough parties to attain the majority. In order to judge whether a coalition will be attractive to a party or not we need to know (or assume) something about that party’s payoffs in case it joins a particular coalition. The table shows the assumed ordinal rankings—first, second, third—for each party over all coalitions. (Thus, smaller numbers are better). These rankings may allow for ties and are based on a combination of the Parties’ statements about their intentions, the assumption that a party will usually prefer being in rather than out of government, some inferences from their observed behavior, and a certain amount of guesswork. They are meant only

to illustrate the concept of dominance and should not be thought of as a research result.

One coalition (such as the grand coalition on the first line) is said to dominate another (such as the improbable one at line 4) if the following conditions hold: Some of the members of the dominated coalition can withdraw from it and enter into the dominant coalition and be better off, and all members of the dominant coalition are better off than they would have been had the dominated coalition been formed.

Let us see if line 1 does indeed dominate line 4. By initiating the shift from the coalition at line 4 to the grand coalition at line 1, the SPD attains a more preferred outcome because they want to honor their promise not to let the Left Party into government. The other member at line 1, the CDU, also attains a more preferred outcome, both because it is in government and because it excludes the Greens and the Left whom it particularly opposes. This shift

makes the Left worse off, both because they are out of government and (probably more important) would prefer to keep the conservative CDU out if possible. However, since the agreement of the Left is not needed to form the grand coalition, their preferences in the matter are irrelevant. Note that while the CSU will of course formally be a member of the Grand Coalition, it is not shown because it is not necessary for a minimal winning coalition. It brings nothing to the coalition, and whatever consideration it receives will be a result of its long-term relationship with the CDU, not because it is needed to form a dominant coalition.

In cooperative game theory, undominated coalitions are stable in a way that dominated coalitions are not; so we may suppose that we are more likely at any particular time to observe undominated than dominated coalitions. It is in just that sense that the coalition at line 4 is said to be improbable.

*Table 8. Minimal Winning Coalitions*

	Minimal Winning Coalitions		Rankings							
				SPD	CDU	CSU	FDP	Green	Left	
1	SPD, CDU	0.65	3	2	2	3	2	9		undominated
2	SPD, Green, Left	0.53	4	8	8	9	1	1		dominated by 1
3	SPD, Green, FDP	0.54	1	6	4	5	3	4		undominated*
4	SPD, Left, FDP	0.55	5	9	8	7	6	2		dominated by 1
5	SPD, CSU, FDP	0.54	2	7	3	5	8	8		dominated by 8
6	SPD, Left, CSU	0.52	7	11	7	8	5	3		dominated by 1
7	SPD, CSU, Green	0.52	6	10	6	6	4	5		dominated by 1
8	CDU, CSU, FDP, Green	0.55	9	1	1	1	6	9		dominated by 2
9	CDU, CSU, FDP, Left	0.56	9	3	5	6	9	6		dominated by 8
10	CDU, CSU, Green, Left	0.54	9	4	6	7	7	6		dominated by 8
11	CDU, FDP, Left, Green	0.56	9	5	9	5	7	4		dominated by 8

The listing in Table 8 includes a number of “unthinkable” coalitions, including several that separate the CSU from the CDU. If these indeed are “unthinkable,” we ought to be able to make that a conclusion, rather than an assumption, of the analysis. Notice, then, that the “unthinkable” coalitions at lines 5, 9, 10 and 11 are all dominated by the coalition at line 8, which the CDU-CSU tried without success to bring about. The one at line 6 is dominated by the grand coalition at line 1. Indeed a coalition including the CSU but not the CDU is improbable on the basis of the game theoretic analysis. Notice also that the coalition at line 8, although dominant over several others, is dominated by line 2, which in turn is dominated by line 1. The coalition at line 2 is “unthinkable” as a formal coalition, since the SPD had pledged not to let the Left into government, but there was some talk about an SPD-Green minority government relying on Left votes. In terms of game theory, this would be a coalition of the three parties – since they would be choosing a common strategy – even though it might not legally be considered a coalition. Notice also that although line 8 is dominated, the coalition that dominates it is itself dominated. There is some controversy in game theory as to whether this sort of transitory dominance should rule out the dominated coalition, in this case the “Jamaican” coalition at line 8. (The parties’ colors, black for the CDU-CSU, Green, and gold for the FDP, correspond to the colors of the Jamaican flag).

We see, also, that the “traffic light” (red, green, gold) coalition at line 3 is, in terms of the definition given, also undominated. However, as the asterisk suggests, there is something odd about it. Line 1 does not dominate line 3 because the SPD, who are a member of the line 1 coalition, are worse off

at that coalition than at line 3. (They would rather be the strongest party in a winning coalition than the second-strongest after CDU-CSU in the grand coalition). However, the Greens and the FDP can both be better off by dropping out of the “traffic light” coalition, since each one prefers to have their traditional bigger partners in government and one another out—even at the cost of themselves being out of government. Thus, they have both the motive and the means to destabilize the “traffic light” coalition by simply refusing to participate in it, and that is what they did. The same motives probably account for the Greens’ destabilizing refusal to participate in the “Jamaican” coalition at line 8.

If “the reasonable objective of governance and public policy is to implement cooperative solutions,” then the study of the cooperative arrangements people do build, the coalitions they form, would seem to play a crucial part in our understanding of governance. (For some instances, see Carraro). This example suggests how, with enough information, we might be able to explain and even predict which coalitions will form and on what terms they will form. An analysis that would do those things is the Holy Grail of cooperative game theory. It would probably also be very useful in a wide range of public policy and governance applications, ranging from models of the membership and form of international organizations to an understanding of the formation of local civil society organizations and their role in local economic development. Unfortunately, many of these probable applications do not quite fit the cooperative-game model, in that the coalitions that form (among sovereign states, for example) are not able to completely control the joint strategies of the members. Nevertheless, even here, a concept

like coalition dominance may be helpful. A comprehensive theory of coalition formation seems to be a very hard problem, but the progress that has been made on it has produced insights that are useful in themselves.

### **Where Do We Stand And Go From Here?**

All of the major approaches to cooperative and noncooperative games incorporate important principles for public policy and have supplied important insights for particular applications. None, however, seems to offer a sufficient framework for policy analysis and the understanding of governance. Noncooperative game theory and bargaining theory, despite some empirical successes, do not take account of the importance of coalitions with large numbers of participants in the real world. Shapley's value theory can always give us a prediction of the payoffs to the members in a coalition, but has little to say about which coalitions will form. The theory of the core can point up some of the difficulties in forming stable coalitions, and suggest that more stable coalition structures are the ones that are more likely to be observed; but may leave open questions about how the benefits gained by a coalition will be distributed among its members.

Ongoing work may generate important new insights for public policy. Some work in cooperative games enriches the theory by letting the payoff possibilities for a coalition depend on the partition of all players into coalitions, rather than just on the membership of the coalition. (Thrall & Lucas 1963, Funaki & Yamoto 1999) Maskin (2004) has extended the Shapley (1953) value approach to include the endogenous formation of coalitions. In his model, some kinds of externalities may persist in a stable coalition structure, with

consequent inefficiency. Maskin's work on this approach is a work in progress, but seems promising both empirically and pragmatically for public policy. As Aumann stressed in his inaugural presidential address to the game theory society, noncooperative and cooperative game theory are two aspects of a single science, and the division of game theory into the two (or more) non-overlapping departments belies that. Research toward a bridge between the two is needed, he suggested. It may be that this can be accomplished only by a middle ground theory that combines aspects of noncooperative and cooperative game theory. In this arena, Brandenburger and Stuart (2004) propose a two-stage game model, in which the first stage is noncooperative, and determines what coalitions form, while the second stage is cooperative. Conversely, to deal with incomplete contracts between employers and employees, McCain (1980) proposed a two-stage game in which the cooperative first stage is constrained by a noncooperative second stage.

Game theory is a young science—not yet quite old enough to be eligible for a Senior Citizen discount on Philadelphia Public Transportation. There is much to be discovered. Indeed much has been discovered. When political economy was a bit older than game theory now is, John Stuart Mill proclaimed that there were no further major discoveries to be made. That was a few years before the emergence of neoclassical economics. Perhaps we will see a similar, and no less useful, flowering in the future of game theory.

### **Selected References**

Aumann, Robert J. (2003) "Presidential Address," *Games and Economic Behavior*, pp. 2-14.

- Barash, David P. (2004) "Caught Between Choices: Personal Gain vs. Public Good", *The Chronicle of Higher Education* Volume 50, Number 32, p. 12.
- Bank of Sweden. (1994) "Prize in Economic Sciences in Memory of Alfred Nobel." <http://nobelprize.org/economics/laureates/1994/index.html>
- Bank of Sweden. (2005) "Prize in Economic Sciences in Memory of Alfred Nobel." <http://nobelprize.org/economics/laureates/2005/index.html>
- Bernheim, B. Douglas, Bezael Peleg and Michael D. Whinston. (1987) "Coalition-Proof Nash Equilibria I. Concepts", *Journal of Economic Theory*, Volume 42, pp. 1-12.
- Brandenburger, Adam and Harborne Stuart. (2004) "Biform Games", Working Paper, NYU and Columbia University, *Harvard NOM Research Paper* No. 00-06, Nov. 2000.
- Burn, Jerome. (1989) *Chronicle of the World*. New York: Prentice Hall Trade Books.
- Carraro, Carlo. (2003) *The Endogenous Formation of Economic Coalitions*. Cheltenham, U. K.: Edward Elgar.
- Dawes, Robyn M. (1980) "Social Dilemmas", *Annual Review of Psychology* Volume 31, pp. 169-193.
- Funaki, Yukihiro and Takehiko Yamato. (1999) "The Core of an Economy with a Common Pool Resource: A Partition Function Form Approach", *International Journal of Game Theory*, Volume 28, pp. 157-171.
- Harsanyi, J.C. and R. Selton. (1987) *A General Theory of Equilibrium Selection in Games*. Cambridge, Mass: MIT Press.
- Macmillan, Margaret (2002) *Paris 1919* New York: Random House.
- Mailath, G. J. (1998) "Do people Play Nash Equilibrium? Lessons from Evolutionary Game Theory", *Journal of Economic Literature* Volume 36, Number 3 pp. 1347-1374.
- Maskin, Eric. (2004) "Bargaining, Coalitions and Externalities", Plenary Lecture, *Second World Congress of the Game Theory Society*, Marseille.
- McCain, Roger A. (1980) "A Theory of Codetermination", *Zeitschrift fur Nationalokonomie*, Volume 40, Number 12, pp. 65-90.
- Mill, John Stuart. (1848) *Principles of Political Economy*. New York: A.M. Kelley, 1987 (Reprint of 1909 edition).
- Nash, John. (1950a) "Equilibrium Points in n-Person Games", *Proceedings of the National Academy of Science*, Volume 36, pp. 48-49.
- Nash, John. (1950b) "The Bargaining Problem", *Econometrica*, Volume 18, pp. 155-162.
- Nash, John. (1951) "Non-Cooperative Games", *Annals of Mathematics*, Volume 2, September, pp. 286-295.
- Nash, John. (1953) "Two-Person Cooperative Games", *Econometrica*, Volume 21, January, pp. 128-140.
- Poundstone, William. (1992) *Prisoner's Dilemma*. New York: Doubleday.
- Rapoport, Anatole and Albert M. Chammah. (1965) *Prisoner's Dilemma*. University of Michigan Press.
- Riker, William H. (1962) *The Theory of Political Coalitions*. New Haven: Yale University Press.
- Serrano, Roberto. (2003) "Fifty Years of the Nash Program, 1953-2003", *Working Paper*, Brown University.
- Shapley, L.S. (1953) "A Value for n-person Games", in H.W. Kuhn and A.W. Tucker (eds) *Contributions to the Theory of Games II*. Princeton: Princeton University Press, pp. 305-317.
- Shubik, M. (1959) "Edgeworth Market Games", in A.W. Tucker and R.D. Luce (eds) *Contributions to the Theory of Games, Volume IV* (Annals of

- Mathematics Studies, 40). Princeton: Princeton University, pp. 267-278.
- Telser, L.G. (1996) "Competition and the Core", *Journal of Political Economy*, Volume 104, Number 1, pp. 85-107.
- Thrall, R. M. and W. F. Lucas. (1963) "n-Person Games in Partition Function Form", *Naval Research Logistics Quarterly*, Volume 10, pp. 281-291.
- von Neumann, John and Oskar Morgenstern. (1944) *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

*Roger Ashton McCain*  
*Department of Economics*  
*Drexel University*  
*Philadelphia, USA*  
*mccainra@drexel.edu*



## Global Development and Financial Institutions

*Richard Peet*

### **Introduction**

The capital investment involved in global development is controlled by big financial institutions. These are of two main kinds: global governance institutions, such as the International Monetary Fund and the World Bank; and private banks, of various kinds, involved in global capital markets. Not only do these institutions supply much of the capital needed to finance development, they also control the financial expertise employed in the investment component that many consider to be crucial to development. Financial institutions of both types, governance and private, consult with each other on a daily basis, form conglomerates to tackle investment-related problems, and regularly exchange personnel and policy documents. They form complexes of financial power exercising widespread control. Banks and other financial institutions see global development as an opportunity for earning interest on deposited funds. Despite attempts, particularly by global governance institutions like the World Bank, but also by private financial institutions like investment banks, to portray development financing as motivated in part by humanitarian concerns, there remains a widespread belief that global finance capital produces a form of development conducive to its own interest – making money. Finance thus becomes a center of controversy, where interests collide over vital issues of global development.

### **Origins of the Institutions**

The idea of establishing international financial institutions to regulate global

development came as a response to growing economic problems in the inter-war years. At first bankers were the main actors rather than governments which, following a liberal ethos, initially tried to avoid direct intervention in explicitly financial international affairs. Far more aware of the effects of national interest rates, for example, in attracting or repelling capital, bankers advocated some kind of international cooperation. For example, the German reparation payments, made under the Dawes Plan of 1924, were negotiated by the chairmen of the boards of the New York Federal Reserve Bank and the Bank of England. There was an international organization in place—the League of Nations—but it remained primarily a “political” organization. However it maintained an Economic and Financial Organization, and it sponsored a series of international conferences in various European cities between 1920 and 1933 (Heilperin 1947; Eckes 1975). In these initial moves we can see the beginnings of global financial institution regulating development.

An international bank to aid post-war reconstruction was first proposed at a conference in Brussels in 1920. Another proposal, to restore the gold standard, stabilized via central bank cooperation, but managed by an international convention, was discussed in Genoa in 1922. In the middle 1920s, the League of Nations helped arrange loans to stabilize the economies of several European countries. An international economic conference, convened by the League at Geneva in 1927, and attended by several non-member countries, such as the US and the Soviet Union, came up with a series of resolutions dealing with trade, cartels and other issues that were thought to constitute an international code of behavior

in policy matters. Discussion of a Bank for International Settlements (BIS) took place in 1930. And since then, regular meetings have been held in Basel, Switzerland, among central bank governors and experts from other financial agencies. The BIS conducts its own research in financial and monetary economics and collects, compiles and disseminates economic and financial statistics, supports the IMF and World Bank, performs traditional banking functions for national central banks (eg gold and foreign exchange transactions), as well as trustee and agency functions. The main international financial institutions, however, were formed by governments through the 1944 Bretton Woods agreement.

The Bretton Woods Institutions were supposed to govern with agreed-upon principles for the conduct of economic affairs decided at the conference. The main principles governing post-war development were as follows:

Principle One. The experience of the competitive depreciation of currencies during the 1930s led to the principle of international control over exchange rates between national currencies under what can be termed “managed flexibility.” Under this, the par values of member country currencies were defined in gold, but allowed to vary around these standards, with par values changed by more than this only under drastic conditions and by permission of the IMF.

Principle Two. A pool of gold and currencies to be drawn on in case of balance of payments difficulties was subscribed by member countries to the IMF according to a quota system adjusted every five years. The quota determined the country’s drawing rights and also its vote in the institution. Countries could exchange specified amounts of their currencies for those of other countries under conditions supervised by the IMF.

Principle Three. To enable multilateral trade, after a five year transition period, all member countries were to eliminate controls making currencies convertible into one another at the official rates without restrictions or discrimination unless approved by the IMF.

Principle Four. Various “scarce currency” provisions were made to decrease instabilities resulting from shortages due to sustained surpluses in a country’s balance of payments—the IMF was authorized to declare the scarcity, ration its remaining supply, require a country to sell currency to it, in exchange for gold, or use other measures.

Principle Five. The Bretton Woods Agreement established a permanent financial institution to promote international monetary cooperation and provide the machinery through which countries could consult and collaborate. The institution, called the IMF, a specialized agency of the UN, was part of an envisaged system that also included a bank dealing with long-term investments (IBRD or International Bank for Reconstruction and Development, later known as the World Bank), a trade organization (ITO), and actions to promote full employment under the UN Economic and Social Council. The IMF would have a board of governors, representing all member countries that met annually, an executive board, of which five members would be from countries with the largest quotas, meeting continuously, and a Managing Director, who was not a board member. Voting on both boards was to be according to quotas. Under the original quota system, the US had 27.9% of the vote and the UK 13.3% (Scammell 1975). The IMF has essentially served as a short term lender to countries in distress mainly over exchange rate depreciations and external debt. The

IMF insists on “conditionalities” to guide a country’s future development before it grants a loan.

The main institution set up to govern development at Bretton Woods was the IBRD. For the first decade of its existence, the IBRD made loans for the post-war reconstruction of Europe. As it turned towards the richer Third World countries in the 1950s, the World Bank concentrated on project loans for infrastructural development. In the late 1960s and 1970s, the World Bank added a basic needs program aimed at poverty alleviation. In 1979 the World Bank began to stress program loans to induce “reforms” in recipient, mainly middle-income countries—by “reform” was meant structural adjustment lending to promote export-orientation and trade liberalization. In this, the World Bank followed the lead of its senior partner, the IMF, under a division of labor that allocated “stabilization programs” (short-term adjustment lending) to the IMF and longer term “structural adjustment lending” aimed at correcting deeper “structural” problems, to the World Bank.

Under this new orientation, poverty took a back seat to new driving forces of macroeconomic policy, stabilization and balance of payments adjustments, all understood within a neoliberal doctrine stressing strict limits to governmental intervention and the virtues of flexible, self-adjusting, free markets. Some commentators find the World Bank shifting at the end of the 1980s and in the early 1990s to a revised neoliberal model stressing market-friendly state intervention and good governance (political pluralism, accountability and the rule of law) with a renewed emphasis on social issues like poverty and education and a dedication to debt reduction. Thus, in the 1990s, the various *World Bank Development*

*Reports* outlined a new “holistic approach” to development involving social safety nets, poverty, health, education, environment, rural areas and gender considerations, in concert with conventionally neoliberal areas, like increased property rights, trade liberalization and privatization. Both the IMF and the World Bank now say that they are committed to global development through debt relief for poor Third World countries and to poverty eradication.

The World Bank and the IMF operate primarily as bankers to the central banks of nation states. They cooperate on a daily basis with finance ministries, with central banks, and with the big private banks. These private banks have power over development policy formation because they control access to capital accumulations through financial markets. Any conception of the regulation of global development by quasi-public financial institutions has to take the broader connection with the banking world into account.

The prominent trade economist Jagdish Bhagwati calls this connection the “Wall Street-Treasury Complex” (in Wade & Veneroso 1998:18). Joseph Stiglitz (2002:230), formerly chief economist at the World Bank says that the IMF in his experience follows “an ideology that was broadly consonant with the interests of the financial community.” Elsewhere the financial connection is described as the Washington-Wall Street Alliance, an institutional complex centered on the US Treasury Department, the IMF and the World Bank, with an intellectual offshoot to Harvard University, particularly the Harvard Business School, whose MBA, doctoral and executive education programs train the corporate and banking elites, but with the leading role being played by Wall Street

bankers, especially investment bankers (Peet *et al* 2003).

This alliance of powerful institutions in the US operates within a broader context of international institutions, such as the Organization for Economic Cooperation and Development (OECD) representing 30 industrial countries, and through organized, inter-governmental meetings. Since 1975, the heads of state or government of the major industrial countries have met annually to deal with major economic and political issues facing the global capitalist system. The most important meeting was the International Conference on Financing for Development, held from 18-22 March 2002 in Monterrey, N.L., Mexico that led to the Monterrey Consensus on financing development.

At first France, the United States, Britain, Germany, Japan and Italy were referred to as the Group of Six (G6) countries, and then with Canada in 1976 the Group of Seven, or G7, and in 1998, with the full participation of Russia, the Group of Eight, or G8. The G7/8 Summit deals with macroeconomic management, international trade, and relations with developing countries – as with debt forgiveness. In addition, the G7/8 holds a series of ministerial meetings among finance ministers, foreign ministers and environmental ministers, among others. Since 1999, the G7/8 has also sometimes met with other “emerging market” countries, such as Argentina, Australia, Brazil, China, India, Mexico, Russia, Saudi Arabia, South Africa, South Korea, and Turkey, referred to as the G20 within the framework of the Bretton Woods institutional system. Even so, a leading issue remains as to why a small number of countries and finance institutions should control development in a world that considers itself to be democratic?

## **Finance Capital**

The term “finance capital” comes from Rudolf Hilferding, an Austrian Marxist theoretician writing in the early twentieth century (Hilferding 1981). The concept referred to the increasing concentration and centralization of capital occurring, in the late nineteenth century, in large corporations, cartels, trusts and banks together with a change in the structure of the capitalist economy towards the export of capital from the industrial countries in search of higher rates of profit. Flows of investment capital served to integrate the nascent global economy predominantly under British control. The term has been revived to refer to the broadly similar globalization of the late twentieth century, this time predominantly under US control.

The US economic system, articulating mass production, mass consumption and mass marketing, proved to be the most successful social, cultural and economic order of the twentieth century. US-based corporations were the most skilled in combining cultural techniques, particularly of market persuasion, with productive techniques, particularly of mass production. US corporations became wealthy beyond their competitors, and Americans became the most affluent people on earth, at least in terms of millions of people with incomes far above the level of basic needs. As a result, one third of global income is earned in the US where the annual GNP amounts to \$10 trillion of a global GNP of about \$31.5 trillion (World Bank 2004).

However this productive power base intensifies to become more exaggeratedly powerful with the move from industrial capital to finance capital. High incomes earned by hundreds of thousands of very rich people and tens of millions of affluent people permit savings, bank accounts,

pension funds, investment and the holding of bonds and shares. Even moving from money to credit increases the importance of the US in finance – so US credit card holders generate 51% of world wide purchases of goods and services (Nilson Report 2003). In terms of the equity (share) market, domestic-listed companies in the US have a market value of \$15 trillion, 49% of global market value of companies listed in all countries, while the annual value of shares traded on US stock markets amounts to three-quarters of the global total trading. In the debt market, all outstanding debt securities issued in the United States amount to about 40% of the global total, while in terms of more exotic financial instruments, 72% of the derivatives (futures plus options) traded on all global markets are traded in North America (BIS 2004).

While financial power accumulates disproportionately in all the leading industrial powers, especially the United Kingdom, Germany and Japan, it is even more disproportionately accumulated in the US. Financial power is centered in a few “global cities” that, in Saskia Sassen’s (1991:4) says, accumulate the capacity for widespread political-economic control by serving as concentrated command points in the organization of the world economy and as key locations for finance and specialized service firms. Control is centered in the few cities that move from being company and corporate headquarters into being “capital markets”, financial centers for national and international economies. The globalization of the late twentieth century saw increased concentration of capital at the top of this global hierarchy, especially in New York, London, Tokyo and Frankfurt, with New York consistently presiding over the rest.

The dominance of New York comes from the amount of capital the city controls and

the technological sophistication with which money can be moved from Wall Street to any where in the world. Essentially this kind of global power is exercised by controlling access to the biggest capital accumulations in the world and directing flows of capital in various forms—as equities, bonds, derivatives, foreign exchange and direct investment—to uses and places that are approved by Wall Street banks and investment firms. Capital markets are filled with financial experts making money from the uncertainties that stem from the normal operation of capitalism – which, after all, is a system whose overall “rationality” emerges from the collision of millions of self-interested actions motivated by profit. Expertise tries to overcome uncertainty by collecting and applying financial and technical information that has to be combined with subjective judgments based on practical experience—expertise is institutionalized experience.

As a result, the annual passage, through a global city like New York, of documents controlling trillions of dollars in commercial and financial assets leaves, as its institutional residue, a system of thousands of specialized companies filled with connections, knowledge and expertise. The combination of expertise, concentrated in specialized institutions, with real control over substantial capital investment, is the basis of the political economic power exercised by global centers of financial power. This power takes many forms.

One form is the influence of capital markets on the making of global development policy by governmental and governance institutions. The most important actors in international capital markets are corporations directly investing in foreign countries and “professional” or “institutional” investors, financial

institutions that invest the savings of individuals, particularly through pension and mutual funds, and invest for non-financial companies, such as insurance companies, in the forms of bonds, equities and loans (Blommestein & Funke 1998). In making these decisions, corporations and investment professionals employ assessments of countries using a range of consultancies and specialized institutions, especially national credit ratings made by agencies like the Economist Intelligence Unit, Business Environment Risk Information, International Country Risk Guide and Euromoney Institutional Investor. Professional investors in such capital markets look at the development policies of governments as signals of the investment risk involved in buying sovereign debt in the form of government bonds.

An assessment that a government bond issue carries a high risk of default or rescheduling, or that the inflation rate will rise, or exchange rates become volatile, means that investors demand a higher interest rate, increasing the cost of debt service, or that bond issues will be under-subscribed or, worse still, withdrawn from the market. The capital market, acting through institutional investment decisions, influences development policy through mechanisms like bond interest rates, direct investment and speculative trading in national currencies.

Some theorists of international capital markets argue that globalization, accompanied by increased market openness and greater capital mobility, has increased the power of capital over government policies (Obstfeld 1998). This is particularly the case for the fiscal and monetary policies of governments that are quickly and efficiently punished by reactions from the international bond and currency markets.

During the 1970s, the financial markets punished governments that persisted in following Social Democratic, Keynesian development policies that investors believed led to state deficits that would impede repayment of borrowed funds (Germain 1997). The paradigmatic case was the disciplining of the British Labour Party in 1976. The terminal case was France's Mitterand government committed to economic expansion combined with income re-distribution, but forced by bond interest rates of 17.4% to reverse course in 1982-3. Since then, policies in developed countries have converged on a neoliberal policy model preferred by the market, consisting of smaller governments, lower deficits, less state provision of social services, lower levels of taxation, less regulation and smaller unions (Dryzek 1996).

Since the middle 1980s international investors have had a high degree of confidence in the policies of what became liberal, rather than social, democracies in the developed world. Financial markets retain a strong influence on government policies but now consider only a limited set of policies in assessing risk premia, leaving some room for policy maneuver and choice (within a basically neoliberal regime) in development policy decisions. By comparison, in the case of "emerging markets" (low or middle income, developing countries) investors retain a strong and broad influence on governmental policy. As Mosley (2003:103) puts it: "although emerging market governments are not always constrained to follow the dictates of global capital markets, the cost of defying [them] is often high".

### **Wall Street**

New York's leading position in global finance capitalism translates into deep and sustained political-economic influence on

development policies. Influence is exerted through a number of power relations linking Wall Street institutions to governments and governance institutions. Standard and Poor's and Moody's, two powerful credit rating institutions, assess the bonds and loans of over 100 sovereign nations, as well as thousands of corporations the world over, to ratify their eligibility for purchase by global investors—this “rating” includes examining the country's economic policy regime—according to “objective criteria” established on Wall Street. A low rating sends a signal to governments that development policies should be changed.

This kind of indirect influence might be termed Wall Street's setting of the “investment climate” within which nation states, corporations, municipalities and others may determine their own policy direction, knowing that some are far more costly than others. Finance capital also directly affects the economic policies of governments through the persuasive influence of Wall Street on the International Financial Institutions (IFIs) involved in global economic governance, especially the IMF and World Bank, institutions that in turn coerce governments forced to borrow outside the private capital markets.

More exactly, the main line of influence between the New York capital market (“Wall Street”) and the IFIs lies through the investment banks. Investment banking is a specialized part of US finance capital created by the Bank Act of 1933, more commonly known as the Glass-Steagall Act. The Act was designed to separate banking from the securities business, and divide commercial banking, taking deposits and issuing short-term loans, from investment banking, concerned more with corporate finance, mergers and acquisitions. Glass-Steagall restricted the primary market,

involving the “floating” of publicly traded corporate securities, to investment banks—“floating” means setting the price and terms on initial public offerings (IPOs) of stocks and bonds that are later traded on secondary markets (such as the stock exchange).

In particular, one of a small number of investment banks, the “special-bracket firms” (traditionally Merrill Lynch, Goldman Sachs, Salomon Brothers, First Boston and Morgan Stanley), takes the lead in forming a larger syndicate of banks that underwrites major issuances of corporate stock and bonds and handles the intricate connections with regulators (as with the Securities and Exchange Commission), the specialized accounting and legal service firms, and the potential buyers in the secondary market. Here the connection with large blocks of capital is crucial because underwriters have to “take up” any unsold shares or bonds in any new issue (Bloch 1986).

Since repeal of Glass-Steagall in 1999, commercial banks, like Chase Manhattan Corporation and Citibank, have developed, or acquired, investment banking facilities, as well as a range of insurance and stock brokerage services, to produce financial conglomerates that are the biggest corporations in the world, making profits even larger than the huge productive and retailing corporations. These financial institutions essentially control access to the New York capital market and its accumulated expertise. Through this they exercise considerable control over global development.

In the post-war period, especially as the neoliberal policy regime was established, many of the key posts in the US government and in the global governance institutions involved with the financial aspects of national and global economic policy were

occupied by experts from the investment banking industry. An elite corps of policy makers circulates between investment banking and the highest levels of governmental and governance institutions. The change in direction of economic policy, from an essentially Keynesian to a neoliberal position, begun during the US Nixon and Ford Republican administrations, solidified in the 1980s, under the Reagan and (first) Bush Republican administrations, and continued in the Clinton and second Bush administrations was overseen by investment bankers (Peet et al 2003).

Investment banking exercises a range of interconnected powers that reach back into the capital market and forward into the production of policy, with control exercised as a combination of expertise, meaning experts armed with specialized knowledge, connections with the financial, corporate and governmental worlds, and command over financial resources, especially large blocks of investment capital. Political-economic power is then exercised on development as combinations of ideas, expertise and finance.

If we re-read the neoliberal “Washington Consensus” development policies, adopted by the financial governance institutions in the late 1980s, and dominant ever since, from the bankers’ perspective, we can see that minimizing state spending, increasing competitiveness, securing property rights (including those of foreign companies), export-orienting economies to produce hard currency, all maximize the loan capacity of developing countries and ensure, to the fullest extent possible, the ability of an economy to repay principle and interest. This does not amount to some kind of cynical conspiracy by bankers and multinational corporations to create puppet economic regimes, despite the evidence that meetings recur among banking and

governmental allies. And once appointed to bureaucratic positions, bankers have considerable freedom to interpret what remains, however, predominantly a banking point of view.

However an export-oriented, privately controlled, market economy *is* the banker’s conception of the “sound” economy, even when re-thought in terms of the more general common good. The problem is that this notion of a “sound economy” does not appear good to those about to be made unemployed, cut off from public services, and paying exorbitant prices for water, electricity and other recently privatized services. The Washington-Wall Street Alliance, the OECD and the G7/8 governments have established, protected and reinforced a neoliberal policy regime that served to de-regulate the world economy (in terms of national state intervention), freeing the way for global, and particularly US corporations, the trading of industrial commodities without interference, and the movement of capital assets across national boundaries that have been reduced in significance.

For critics, rather than a “sound” global system, the result is a wild economy of colliding interests and immanent debt crises beyond the control of any particular interest or institution. It is evident, from bankers’ emphasis on financial reform in debtor countries, and in their belief that real economic growth is the only way debtor countries can service debts and repay loans, that they think the financial structure and the quality of financial services play a decisive role in making real growth possible, and that, reciprocally and mutually, investment in production and in long term development projects is crucial to a country’s financial status, its creditworthiness with regard to banks, and by extrapolation, to the



international financial system and its stability. The IMF and the World Bank were established in 1944 to play this dual role, of stabilizing the global financial structure and encouraging economic development. The questions remain, however, what kind of development, towards what ends, and in whose interest?

The most difficult question of all in the area of international finance is control over development by “The Market”, meaning the various capital markets centered on Wall Street, New York City. Saying that Wall Street is the most powerful actor in setting the policies used by global governance institutions to develop countries the world over implies a more conscious collective actor than may indeed be the case. The central financial institutions located on Wall Street are the equity and bond markets, and these no-one controls. While financial analysts may objectively “study the numbers” while balancing return with risk in evaluating whether to invest client money in the bonds or shares issued in foreign countries, the decision is rendered into subjective judgment by uncertainty about the future when the bonds will be re-paid or the shares sold again. George Soros, head of the Quantum Fund, “the best performing investment fund in history,” says of financial markets:

Each market participant is faced with the task of putting a present value on a future course of events, but that course is contingent on the present values that all market participants taken together attribute to it. That is why market participants are obliged to rely on an element of judgment. The important feature of bias is that it is not purely passive: It affects the course of events that it is supposed to reflect. This active ingredient is missing from the concept of

equilibrium employed by economic theory (Soros 1998: 47).

The international financial market, then, is a meeting point where biases and guesses collide. Yet the collision of uncertainties in the market determines the economic fate of nations, states and the global economy. As one commentator says: “National governments find themselves in environments not merely of risk but of radical uncertainty .... governments often cannot know whether the response of world markets to their policies will be merely to make them costly or to render them completely unworkable (Gray 1998:74-5). Global development, directed by international financial institutions, private and public, has this radical uncertainty at its heart.

### **Internet Sites**

BIS (Bank for International Settlements). [www.bis.org/statistics/secstats.htm](http://www.bis.org/statistics/secstats.htm)  
International Conference for Financing for Development. [www.un.org/esa/ffd](http://www.un.org/esa/ffd)  
International Financial Institutions Research Sites. [www.wellesley.edu/Economics/IFI/](http://www.wellesley.edu/Economics/IFI/)  
International Monetary Fund. [www.imf.org/](http://www.imf.org/)  
World Bank. [www.worldbank.org](http://www.worldbank.org)

### **Selected References**

Bloch, E. (1986) *Inside Investment Banking*. Homewood: Dow Jones-Irwin  
Blommestein, H.J. and N. Funke. (1998) (Editors) *Institutional Investors in the New Financial Landscape*. Paris: OECD  
Dryzek, J.S. (1996) *Democracy in Capitalist Times*. New York: Oxford University Press  
Eckes, A.E., Jr. (1975) *A Search for Solvency: Bretton Woods and the International Monetary System, 1941-1971*. Austin: University of Texas Press.

- Germain, R. (1997) *The International Organization of Credit: State and Global Finance in the World Economy*. Cambridge: Cambridge University Press
- Gray, J. (1998) *False Dawn*. New York: The New Press
- Heilperin, M.A. (1947) *The Trade of Nations* New York: Knopf
- Hilferding, R. (1910) *Finance Capital* London: Routledge and Kegan Paul, 1981.
- Kaufman, G.G. (1986) *The US Financial System: Money, Markets, and Institutions*. Englewood Cliffs: Prentice Hall.
- Kidwell, D.S. and Richard L. Peterson. (1987) *Financial Institutions, Markets, and Money*. Chicago: The Dryden Press
- Mosley, L. (2003) *Global Capital and National Governments*. Cambridge: Cambridge University Press
- Nilson Report. (2003) [www.nilsonreport.com](http://www.nilsonreport.com)
- Obstfeld, M. (1998) "The Global Capital Market: Benefactor or Menace?", *The Journal of Economic Perspectives*, Volume 12, pp. 9-30
- Peet, R. et al (2003) *Unholy Trinity: The IMF, World Bank and WTO*. London: Zed Press
- Sassen, S. (1991) *The Global City: New York, London, Tokyo*. Princeton: Princeton University Press
- Scammell, W.M. (1980) *The International Economy Since 1945*. New York: St. Martin's Press
- Sorros, G. (1998) *The Crisis of Global Capitalism*. New York: Public Affairs
- Stiglitz, J. (2002) *Globalization and its Discontents*. New York: W.W. Norton
- Wade, R. and F. Veneroso. (1998) "The Asian Crisis: The High Debt Model versus the Wall Street-Treasury-IMF Complex", *New Left Review*, Number 228, pp. 3-23
- World Bank. (2004) *World Development Report*. New York: Oxford University Press

*Richard Peet*  
*Graduate School of Geography*  
*Clark University*  
*Worcester, Massachusetts, USA*  
*rpeet@clarku.edu*

### **Introduction**

For the purpose of this article, global policy coordination refers to economic coordination that involves information exchange, dialogue, common objectives, joint action, and the endorsement of current policies (Truman 2004). Most researchers point to externalities as a fundamental rationale for policy coordination (Hagen and Mundschenk 2001; Barrell, *et al* 2000). When national economies are interdependent, the policy actions of one country may generate spillovers in other national economies. Policy coordination may internalize these externalities to achieve higher welfare when countries take these effects into consideration. Some (Diaz-Roldan 2004; Frankel *et al* 1990) also argue for the need for policy coordination from a "public goods" point of view.

Public goods such as open markets, exchange rate stability, a common market, a single currency, etc. benefit all parties involved. In the context of an economic community or monetary union, public goods take the form of "club goods" (Hagen and Mundschenk 2001). Another benefit of policy coordination is mutual education within macroeconomic policymaking which one national authority cannot accomplish alone (Gagnon 2004). Through policy coordination, national governments are able to learn something about their policies and the need for improvement. For example, the message regarding the need for restructuring the Japanese banking system and assets of its balance sheet has been reinforced to Japanese authorities through the G-7 summits since the 1990s (Truman 2004).

Policy coordination as a research topic has drawn attention in recent decades due to increasing interdependence in world economy. Global policy coordination can occur in areas such as macroeconomic policy, exchange rates, international trade, and the debt crisis of developing countries. Since many of these issues are extensively covered in other articles in this publication, this article emphasizes macroeconomic policy and exchange rate policy coordination, the two areas of policy coordination that are highly visible and politically sensitive in the national economic and political debates. This article begins with a discussion of the history of policy coordination, followed by a discussion of macroeconomic policy coordination and the related exchange rate stability regimes. The article then discusses macroeconomic policy coordination and coordination in other policy areas, followed by a discussion of the political economy of policy coordination.

### **History of Policy Coordination**

During the 19<sup>th</sup> century, policy coordination took the form of monetary and exchange rate policy coordination by central banks (Fischer 1988). Central banks in major European countries lent to each other to stem the balance of payment crises. Some countries went so far as to establish monetary unions such as the Latin Monetary Union which joined Belgium, Italy, and Switzerland with France, and the Scandinavian Monetary Union which included Sweden, Denmark and Norway (Cohen 1998). Member states in a monetary union accepted each other's currencies, even banknotes and drafts. Maintenance of the gold standard and gold convertibility imposed discipline on the monetary policies of nations. To correct external disequilibrium, countries, instead of

allowing gold to float across borders as dictated by Hume's price-species flow mechanism, adopted tight money to stem gold flow and defend gold convertibility (Bloomfield 1959). Slow growth was the cost paid for such a policy.

During the interwar era of the 20<sup>th</sup> century, countries made efforts to return to the gold standard but faced the same dilemma as in the 19<sup>th</sup> century, as the need for external balance required non-expansionary domestic monetary policy and frustrated the countries' effort to increase employment (Fischer 1988). Britain returned to the gold standard with the overvalued pound. After the Great Depression, when small countries that held pounds as reserves lost confidence in sterling-gold convertibility and began converting pound for gold, Britain was off gold in 1931 (Eichengreen & Jeanne 2000). The United States and France devalued their currencies in 1935 and 1936 respectively but remained on the gold standard. Countries that clung to the gold standard without devaluation suffered the most during the Great Depression. There were tremendous reserve movements for countries that maintained pegged exchange rate systems, while there were sharp exchange rate fluctuations for countries with floating exchange rate regimes (Krugman & Obstfeld 1997). Some governments also adopted capital control and administrative allocation of foreign exchanges. All these were just predecessors to worldwide fixed exchange rate systems from 1945 to 1973 and floating exchange rate regimes since 1973.

The Bretton Woods agreement in 1944 established a form of policy coordination based on an exchange union among major industrialized countries (Cohen 1998; Webb 1995). It adopted fixed, but adjustable, exchange rates anchored by the dollar which

was pegged to gold. Although it provided nearly a quarter century of currency stability, maintaining fixed exchange rates imposed a constraint on domestic policies, the objectives of which may have been inconsistent with external balance (Baines 2001). Although the International Monetary Fund (IMF) was expected to lend to countries to address short-term external imbalances, countries were expected to alter or coordinate their domestic policies in the long run. For example, from 1964 to 1966, Britain adopted restrictive policy, in conjunction with international borrowing, to maintain the pound's value against the dollar. However, more frequently, countries were not willing to give up their domestic policies to defend fixed exchange rates. For example, fiscal expansion in the second half of the 1960s in the United States, accommodated by expansionary monetary policy in 1967 and 1968 pushed up inflation, contributing to the sharp deterioration of the U.S. current account balance in the early 1970s. With increasing capital mobility, balance of payment disequilibrium became more frequent and significant, which triggered further speculative capital flow and built up more pressures for exchange rates to adjust. In the end, the government-engineered or coordinated exchange rate adjustments simply became too slow and too cumbersome. The collapse of the Bretton Woods agreement in the early 1970s, after a series of balance of payment crises, was a natural outcome of the correction for the exchange rate misalignment (Krugman & Obstfeld 1997).

In the era after the end of Bretton Woods, it has become increasingly evident that a benefit of flexible exchange rates touted by proponents, the ability to insulate the domestic economy from foreign disturbances, has failed to materialize

(Lowell 1994). There have been hard peg and the effort toward exchange unions, flexible exchange rates, and the intermediate exchange rates (Bofinger & Wollmershauser 2003). Although the general trend has been movements toward two corner solutions against the intermediate exchange rates, studies have revealed that some independent floaters are actually engaged in various interventions such as smoothing, targeting, or managing their floating exchange rates (Bofinger & Wollmershauser 2003). As a result, stabilizing exchange rates, or altering their course, has always been at the core of international economic coordination. Coordination during the floating exchange rate era has taken the form of the ad hoc Group of Seven (G-7) process and regional monetary integration.

The G-7 process involves annual summits participated in by the seven major industrialized countries (originally G-5) along with meetings of their finance ministers and central bankers, and finance deputies and central bank deputies (Meyer *et al* 2002). During the 1970s, coordination took the form of joint expansion as reflected in the "locomotive effect," embodied in the 1977 London Summit, and especially in the 1978 Bonn Summit when Germany, France, and Japan pledged fiscal expansion targets, and the U.S. agreed to decontrol its oil prices. The "triumph of monetarism" from 1980 to 1984 fostered decentralized policy making and diminished coordinated actions. From 1985, pragmatism prevailed in policymaking, leading to major foreign exchange rate policy coordination, as enshrined in the Plaza Accord of 1985, which, in the main, aimed at reducing pressures on an appreciating dollar as a result of fiscal expansion and tight money in the United States (Frankel 1988). Other targets included the reduction of the U.S.

budget deficit and West Germany fiscal expansion. The 1987 Louver Accord concluded that the dollar depreciation had gone far enough, and targeted a continuous reduction of the U.S. fiscal deficit, a Japanese monetary expansion, among other things. In 1989, there were additional coordinated policy actions to prevent the dollar from rising again. Since the 1990s, coordinated policy actions have declined significantly. A few episodes of coordination include interventions in 1994 and 1995, and again in 1998, to prevent the yen from declining. However, increasingly, the G-7 process has become an occasion for dialogue and information sharing, instead of setting tangible policy targets for countries to follow (Meyer *et al* 2002). During recent years, with Russia's weight in the world's oil and natural gas production and export, and the rising economic clout of developing countries such as China, India, South Africa, and Brazil, representatives from these countries occasionally are invited to attend some G-7 side meetings. In February 2006, Russia for the first time holds the presidency of the G-8 meeting, which devotes its attention to world energy issues, among others. The G-7 or G-8 summit meetings increasingly become a forum of exchange of information for major players in the world economy in an effort to maintain a sustained balance in the world economy (Financial Times 2006a).

At the regional level, a group of European countries joined together in the 1970s in an effort to stabilize their currencies with respect to each other, leading to several European exchange rate management systems such as the "snake", "mini-snake", and the "snake-in-a-tunnel." All these strategies involved a few core countries that pegged their currencies within a certain margin of each other but jointly

floated against other currencies (Krugman & Obstfeld 1997). In the late 1970s, the European Monetary System or EMS was established among members of the European Economic Community. The EMS adopted the Exchange Rate Mechanism (ERM) which established a central exchange rate between the European Currency Unit (ECU) and outside currencies. The ECU is the European unit of account, the value of which was determined as a weighted average of the participating currencies. The value of member currencies was kept within a certain margin of the central exchange rate. The numerous realignments of the ERM and its breakdown in 1992 proved that the system was not insusceptible to outside disturbances. In 1999, the EMS evolved into European Monetary Union with 12 members from the EU as member states (Hagen and Mundschenk 2003). Slovenia joined in 2007 as the 13<sup>th</sup> member. The monetary union adopts the euro as the single currency and a common monetary policy that is conducted by the European Central Bank (ECB). Member countries are required to limit their fiscal deficits to within 3% of their GDP as reflected in the Growth and Stability Pack, though major countries such as Germany, France, and Italy repeatedly breached the threshold, causing the eventual collapse of the fiscal pack in 2004. Nonetheless, the evolution from the snake peg to the monetary union reflects an increasing degree of macroeconomic and foreign exchange rate policy coordination among the core European Union countries.

There are a few other operating regional monetary unions. The CFA Franc Zone in Africa has brought a number of African countries into two monetary blocs, the West African Economic and Monetary Union and the Economic and Monetary Community for Central Africa. Each bloc issued its own

legal tender pegged to the French franc. The moneys were jointly managed under the aegis of the French Minister of Finance as integral parts of a single currency (Cohen 1998; Masson & Pattilo 2004). In the years following independence, the CFA bloc as a whole had lower inflation rates and higher growth than other non-member sub-Saharan African countries. However, since the mid 1980s, the appreciating franc brought stress to the export sector of the bloc resulting in current account and budgetary imbalances, and capital flight. In 1994, after 50 years of fixed exchange rate with the French franc, the two CFA franc zone moneys devaluated against the franc, which, along with rising raw material prices, helped bring down inflation and facilitate trade growth. Since 1999, the bloc switched to peg its currencies to the euro. An appreciating euro in recent years, along with rising oil prices and declining agricultural prices, has brought uncertainty to the bloc's economic prospects.

In southern Africa, the Common Monetary Union consists of South Africa and small countries such as Lesotho, Swaziland, and Namibia, which initially adopted South Africa's rand but later only pegged their national currencies to the rand. Such an exchange rate union has generally benefited from low inflation and there is the evidence of per capita income convergence in the union. These African monetary unions, along with unsuccessful monetary integration attempts from the 1920s to 1960s among some Eastern African countries under the rubric of the Eastern African Community and Common Market, have precipitated the search for an African monetary union by some. The desire is further fueled by the success of the European Monetary Union. In 2003, the Association of African Central Bank Governors announced the plan for a single

currency and common central bank for the continent in 2021 (Masson & Pattilo 2004). However, research has revealed the difficulties ahead due to drastic asymmetries among countries in terms of size, economic structure, fiscal discipline, and macroeconomic stability. A single currency alone is hardly believed capable of delivering low inflation and stable exchange rates, as it intends to (Masson and Pattilo 2004). Similar asymmetric problems are also found to be a hindrance to the plan for a monetary reunion between Russia and Belarus (Chaplygin 2004). On the other hand, the planned monetary union with a single currency pegged to the dollar, among members of the Cooperation Council of the Arab States of the Gulf in 2020, may not significantly affect the actual operation of macroeconomic policy of the region since most member states are currently already pegging their domestic currencies to the dollar (Igbal 2002).

The successful launch of the European Monetary Union also fueled academic interest in monetary unions in Latin America and Asia. Ferrando (2000) examines the feasibility of the North America Free Trade Agreement (NAFTA) and the Mercoser as separate monetary unions. His primary finding is that compared with the EU and the NAFTA, the Mercoser has had a much poorer record in terms of unidentifiable trade correlation schemes, uneven sizes among member states, low economic interdependence within the bloc, high variability of exchange rates, poor track record of inflation, low production diversification, etc. Therefore, the Mercoser's feasibility as a monetary union is far from that for NAFTA. Similarly, Temprano-Arroyo (2003), based on Optimal Currency Area theory, concludes that none of the regional trade agreements in Latin

America are suitable for monetary integration.

On the other hand, Branson and Healy (2005) find that the Association of South-East Asian Nations and China (ASEAN plus 1) have similar trade patterns. They suggest that there has been implicit policy coordination among ASEAN countries, and even among the ASEAN countries and China. For example, exchange rates among these countries show strong positive correlations. Monthly changes in reserves among these countries are also positively correlated. They argue that ASEAN and China are qualified for explicit exchange coordination in the form of a common currency basket (Branson and Healy 2005). Indeed, there has been an ascendance in economic cooperation in recent years between ASEAN and East Asian countries within trade and investment. There has also been early stage of monetary coordination in the form of the Chiang Mai Initiative, a bilateral swap arrangement among the ASEAN plus 3 major East Asia countries – Japan, China, and South Korea (Kawai & Takagi 2005). In addition, there has been increasing levels of information and exchange. Asian Development Bank launches a notional unit of exchange called Asian Currency Unit (ACU) in early 2006. The ACU is modeled on the ECU and is designed to track the relative values of Asian currencies. The ACU will be calculated using a basket of 13 regional currencies, weighted according to the size of each economy (Financial Times 2006b). Whether this will help facilitate actual macroeconomic coordination remains to be seen. Interestingly, Fernandez-Arias and Montiel (2001) examine why growth of Latin American countries lagged behind that of East Asian countries for the 1990s. They find that less complete macroeconomic



reform in Latin America than in East Asia indeed accounts for half of the growth gap between the two regions. The other half is accounted for by the external environment, structural and institutional reforms, and the level of education of the labor force. Their conclusions suggest that macroeconomic policy and macroeconomic coordination matter, but that they are not everything. Some scholars propose a three exchange rate bloc regime with a world trading system centered on the European Union, North America, and Japan-dominated East Asia. According to this idea, within each trading bloc, countries closely coordinate their exchange rate policy through fixing their currencies against each other (i.e. the EMS) or even adopting a regional currency (i.e. the euro). Currencies of different trading blocs float against each other. Although the three-currency bloc remains a hypothesis, developments in Europe have raised the hope that this model may one day serve as the mechanism for global-wide exchange rate coordination (Cohen 1998; Lowell 1994).

### **Coordination of Macroeconomic Policy**

Macroeconomic policy coordination refers to the efforts of national governments regarding their fiscal or monetary policies in the form of dialogue, sharing information, policy targeting, or joint policy actions designed to improve macroeconomic performance. The macroeconomic issues facing a national government are generally those of output, employment, inflation, and external balance. The instruments at the disposal of governments are mainly fiscal and monetary policies (Schultze 1988). When using these tools to nudge the economy toward a desirable state, the working of the policies and the economy may be in conflict with other economies that

are also being nudged by their own authorities. For example, a tightening of monetary policy in one country to reduce inflation will result in an appreciation of the exchange rate which can adversely affect the inflation prospects of other countries as imports become more expensive (Barrell *et al.* 2000). Policy conflicts (temporary or stabilization conflicts) are conceivable when national economies are at different stages of business cycles, with one being at full employment and the other in a recession. Thus, expansionary policy would push the former into an inflationary path and help initiate growth in the latter. Permanent policy conflicts (ongoing conflicts) come as a result of conflicting objectives (Meyer *et al.* 2002) with governments preferring different levels of budget deficits or inflation targets. In an interdependent world economy, policy conflicts give rise to divergent welfare effects. Policy coordination seeks to redistribute welfare benefits so that all sides will remain on sustainable paths.

Early studies of policy coordination, or the “first generation models,” adopt the Mundell-Fleming model of open macroeconomics, which provides a stylized mechanism of the interactions involved. Under fixed exchange rate regimes, monetary expansion will cause rising income and employment, leading to a deteriorating current account. Reserves (when capital is immobile) or both reserves and capital (when capital is mobile) will flow out of the country to finance the rising current account deficit and to induce a lower price to improve the current account balance. The outflow of reserves cancels out the effect of loose money in the long-run. On the other hand, fiscal expansion may crowd out private investment through higher interest rates and maintain the same level of



income and employment (when capital is immobile) or lead to rising income and employment and the inflow of capital to finance the current account imbalance (when capital is mobile). Alternatively, under flexible exchange rate regimes, income expansion caused by loose monetary policy will be followed by a deteriorating current account which may be balanced by the depreciating currency (when capital is immobile). When capital is mobile, capital outflow and even greater currency depreciation will balance the current account more quickly. On the other hand, fiscal expansion will lead to capital inflow and currency appreciation, offsetting the initial income and employment increase (Caves *et al.* 1999).

The above stylized transmission mechanism illustrates the crucial elements for policy coordination under alternative exchange rate regimes. Since a country's income growth will cause deterioration of its own current account, the improvement of a country's external balance will largely depend upon its trading partner's willingness to expand its economy. A system that maintains a fixed exchange rate among trading partners requires coordinated domestic policymaking. Countries with balance of payment surplus and accumulating reserves should expand while those with balance of payment deficits and declining reserves should contract in order to stabilize the exchange rate. In particular, countries must refrain from inflationary policy in order to avoid depreciating pressures on any currencies. Thus, the fixed exchange rate system imposes a discipline on participating countries, and governments lose independence in their monetary policy. The goal of internal balance (full employment and inflation) may contradict that of external balance (balance of

payment). Although capital mobility provides countries with means to finance their current account deficits, the prospect that foreign investors will stop purchasing deficit countries' assets may put stress on the exchange rate, and even trigger speculative attacks. This explains the collapse of the Bretton Woods system: a lack of coordination of domestic policies and the related exchange rate misalignment and associated speculation on dollar devaluation.

Under flexible exchange rates, countries are free to pursue their domestic policy goals hoping the floating exchange rates will shelter the domestic economy from external shocks. Monetary authorities tend to seek price stability by altering money supply, causing the price of currencies to swing widely, resulting in exchange rate volatility. Contrary to the original claim that flexible exchange rates will insulate domestic economies from external shocks (Mussa 1979; Coeure and Pisani-Ferry 1999), the wide swings in exchange rates may have significant impacts on the domestic economy. Significant and persistent exchange rate misalignment will eventually become a policy concern since it may involve sharp currency depreciation, inflation, and high interest rates, which eventually may bring growth to a halt. This explains a major recurrent theme within policy coordination under flexible exchange rates: concerns about government's domestic agenda and exchange rate realignment. Many of the academic studies on policy coordination during the flexible exchange era focus on stabilizing exchange rates.

Some recent studies of policy coordination adopt the New Open Economy Macroeconomics, or "second generation models." The new models are designed to address open economy issues in a dynamic general equilibrium version of the open

economy with nominal rigidities and market imperfections. The new models are built on microfoundations and seek intertemporal maximization, instead of merely tracing interactive effects of macroeconomic parameters in a static environment, as in the Mundell-Fleming models. In particular, with market power for agents, the new models are believed to enable provision of an explicit justification for the Keynesian assumption of short-run demand determined output. Furthermore, with utility maximizing agents, the new models can be used to analyze the welfare implications of policy decisions within the context of the model. While the Mundell-Fleming model continues to be the dominant model of open economy macroeconomics, the new models yield new insights into the issue of policy coordination (Obstfeld and Rogoff 2000; Canzoneri et al. 2004). Corsetti and Pesenti (2001) present a variant of the new model that contains monopolistic firm level production functions with differentiated goods, labor demand functions, and consumer utility functions, in addition to fiscal and monetary rules. Firms' markups are exposed to exchange rate fluctuations. Without coordination, an inward-looking monetary policy may generate excessive exchange rate volatility, which leads exporters to set higher prices in response to high profit risk. The higher import prices lead to a large domestic output gap. Monetary policy coordination may reduce exchange rate fluctuation, and reduce import prices and thus the output gap (Corsetti & Pesenti 2001). However, Obstfeld and Rogoff (2002) find that while coordination is beneficial in theory, it may be relatively unimportant empirically. This conclusion is further confirmed by Sutherland (2002).

Macroeconomic policy coordination also depends on the national governments'

strategic reaction to effects of economic interdependence (Diaz-Roldan 2004). Assume a transmission mechanism in a two-country world where an expansionary policy in one country will cause its current account to deteriorate and growth in the GDP of its trading partner. Also assume that national authorities' objectives are price stability and balance-of-payment equilibrium. A game theoretic framework, first adopted by Hamada (1976) and then widely used in the policy coordination literature, illustrates incentives for potential gains from policy coordination or non-coordination. If a country assumes the policy of its trading partner as given, an expansionary policy would leave it worse off (deteriorating current account) and the partner country better off (GDP growth and full employment). As a result, the country will choose not to expand in order to avoid this non-cooperative or the Nash equilibrium. For the same reason, the partner will do the same. The end result will be that both countries choose not to expand and thus both suffer unemployment. However, if both countries coordinate their expansion so that their economies grow together, neither country will suffer significant current account deterioration and both will achieve full employment. Under this coordinated equilibrium, an individual country may gain less than when only one country expands. However, it avoids the worst scenario where both countries do not expand (Fischer 1988). There is a large literature examining strategic coordination. The general conclusions are that if authorities ignore interdependence, the solution would not be optimal; but if authorities coordinate, the result would be Pareto superior (Gutierrez 2003).

The nature of macroeconomic transmission also helps shape the strategy of

policy coordination. Macroeconomic transmission can be positive where expansion in one country produces expansion and an improved current account in another. The 1978 German fiscal expansion was intended to have such an effect in other countries where the German economy would act as a “locomotive” to bring growth to its trading partners through expanding exports to Germany. Macroeconomic transmission can be negative. For example, a monetary expansion in one country may cause higher real interest rates and thus contraction in another. Transmission can also be asymmetrical where monetary expansion from one country generates a more powerful effect in its trading partners than the other way around. The nature of policy transmission may help determine optimal policy coordination. For example, when there exists a negative transmission from a monetary expansion, the optimal policy coordination should be less monetary expansion in both countries. For positive transmission, a worldwide disturbance calls for similar policy responses in different countries.

Although there is strong theoretical basis for policy coordination, many issues can defy the general principle and generate inferior outcomes from coordination. Frankel (1988) points out three obstacles to coordination: compliance, credibility, and uncertainty. In terms of the compliance problem, countries have the incentive to cheat and thus gain when assuming the policies of other countries as given. This is likely to happen when the commitment to coordination cannot be unambiguously monitored, especially when the potential indicators to be monitored are many. As a partial solution, Meyer *et al* (2002) suggest that a supranational authority will more

likely make countries follow efficient policies since it can punish “defectors.” Examples include the European Monetary Union and the European Central Bank. In addition, repeated coordination gives incentives for countries to stick to their commitment through a coordination agreement instead of being a free rider or breaking rank. The credibility problem occurs when countries commit to joint expansion, which eventually leads to inflation (Frankel 1988). Uncertainty arises when countries are not sure of their economies’ position, the most suitable policy targets, or the response of the economies to policy changes. There are other uncertainties. Alternative models of simultaneous equation estimation of interactive macroeconomic policy impacts (Caves *et al.* 1999; Fischer 1988) in the United States and the rest of the Organization of Economic Cooperation and Development (OECD) demonstrate a high degree of divergent effects of monetary policy. These inconsistent and divergent results raise the question of “model uncertainty” (Meyer *et al.* 2002). There is no certainty that models actually capture the essence of national economies or that the parameters reflect true underlying adjustment processes. Countries may have different views about how the economy works and what the effect of policy is. In addition, countries may not know what the true underlying model is. As a result, coordination to achieve an agreement may result in a worse outcome than that from following a non-cooperative strategy (Frankel *et al.* 1992).

Although theoretical arguments for policy coordination are strong, most empirical studies on the size of actual benefits from policy coordination find relatively small gains from policy

coordination (Caporale, *et al.* 2003). Although the actual estimation varies, studies generally show gains from coordination are usually at a magnitude of between one-half and one percent of GDP per year (Oudiz and Sachs 1984; McKibbin 1997). Similarly, small gains from policy coordination are found in Hallett (1986) and Currie *et al.* (1987, 1989). Two approaches are generally adopted in estimation. The policy optimization approach measures the size of potential gains generated by moving from a Nash non-cooperative game to Nash bargaining (Sachs and McKibbin 1985; Ishii *et al.* 1985). Another approach evaluates the welfare effects of alternative international regimes, which are characterized in terms of policy rules (Currie & Wren-Lewis 1990; Williamson & Miller 1987).

Diaz-Roldan (2004) sees monetary integration as one of an infinite number of ways toward macroeconomic policy coordination. Bryant (1995) develops a macroeconomic policy coordination typology, which is based on two dimensions: the different degrees of use of cross-country agreements and international institutions on the one hand, and different degrees of activism in macroeconomic policy interference on the other. The different degrees of macroeconomic policy interference range from a non-activist approach with simple rigid rules, to the use of rules of limited scope for discretionary adjustment, to moderately activist discretions but with some announced constraints, to highly activist discretion in use fiscal and monetary policies. Based on this scheme of topology, Bryant classifies the gold standard of the late 19<sup>th</sup> century and early 20<sup>th</sup> century as combining low cross-country agreement, very weak use of international institutions, and less than active activist discretionary use of macroeconomic

policy. According to Bryant's scheme, all episodes of late 20<sup>th</sup> century macroeconomic policy coordination, including Bretton Woods in the 1960s, post-Bretton Woods from 1971 to 1985, and the recent past from 1985 to the early 1990s, involve much higher activist discretionary decisions. Interestingly, Bryant classifies many of the proposed reforms in international economic policy coordination as types of coordination where there are much higher levels of involvement in cross-country agreement and international institutions.

### **Exchange Rate, Other Policy Coordination**

Macroeconomic policy and exchange rate policy are among the many tools at the disposal of national authorities to manage the economy with the goal of internal and external economic equilibrium. Examples of other policies include structural policies (labor market, industry regulation, etc.), financial market policies (capital control, banking, etc.), trade policy, and population and immigration policies. These policies affect resource allocation at least just as effectively as fiscal, monetary or exchange rate policies. Thus, there may be trade-offs between macroeconomic-exchange rate policy and other policies. In many cases, these policies influence and compliment each other (De Grauwe & Mongelli 2005; Ferrando 2000). For example, capital control policy affects capital mobility and thus has direct impacts on how macroeconomic policy and exchange rate policy work, as the Mundell-Fleming model suggests. Lessoning capital control may induce interest rates to converge among countries and thus reduce the need for monetary policy to stimulate investment. Labor market reforms and industrial deregulation may help reduce distortion and increase structural flexibility,

and thus reduce the need for monetary or fiscal policy directed toward increasing domestic employment and income. Full employment can be achieved through cross-border movement as well as cross-sector movement in domestic markets. Mobility of the labor force helps reduce the burden on macroeconomic policy in balance of payments imbalance adjustment since demand for imports can be lessened through out-migration instead of a recession (Krugman & Obstfeld 1997). Both trade and factor movements can help cross-country price equalization, and thus reduce the likelihood and extent of inflation shock (Rogoff 2003).

Price and structural flexibility is at the heart of Optimal Currency Theory by Mundell (1961). When countries give up their monetary policies to form a monetary union, they trade their monetary policies (even fiscal policies) off for economic gains associated with a currency union such as price stability, exchange rate stability, a credible central bank, etc. (Hagen and Mundschenk 2003). Any imbalances of the sub-units within the union will have to be corrected through micro-level adjustments. States in the United States do not have their own monetary policy, and mostly have a mandate to balance their budget. The lack of state-specific monetary and fiscal policies (in other words, there is uniformity of monetary and fiscal policies across states) is compensated for by flexible price and factor markets, and by the federal income transfer system (Krugman & Obstfeld 1997). This is a case of complete substitution of microeconomic adjustments for macroeconomic and exchange rate adjustment. In the euro zone, countries give up their monetary policies but retain certain fiscal autonomy. In addition, cultural differences and structural rigidities in factor

markets and in the service sector among member states have prevented microeconomic adjustments from eliminating all sub-union imbalances. High unemployment and slow growth in some major euro countries indicate that the benefits of macroeconomic coordination will be limited without flexibly adjustable microeconomic structure. Important goals of the Lisbon Agenda since 2000 are to institute domestic regulation reforms and labor markets and pension reforms, along with a reorientation of public spending, in order to increase the flexibility and competitiveness of the euro zone economies (Anonymous 2005). Microeconomic flexibility reduces the extent of the required macroeconomic adjustment (and thus the extent of macroeconomic coordination) since spillover effects can be absorbed through price changes and/or factor movement (De Grauwe & Mongelli 2005; Ferrando 2000).

However, certain microeconomic adjustments may face stiff domestic resistance and thus carry high political risks. This may make macroeconomic and exchange rate adjustment and coordination the first choice of politicians since foreigners may share part of the cost of adjustment (Meyer *et al.* 2002; Fischer 1988). Political officials choose policies that are politically convenient instead of economically sound. For example, leaders from Germany and France suggest a European Union wide “tax harmonization” to promote common corporate tax policies. This would stem the capital mobility that the European monetary union intends to promote. Political officials also tend to use international macroeconomic efficiency to compensate for domestic microeconomic inefficiencies.

Indeed, most discussions of macroeconomic and exchange rate

adjustment and coordination implicitly assume away politically sensitive issues such as immigration and domestic reform. Permitting only limited changes in microeconomic factors limits the choice of government and forces macroeconomic and exchange rate adjustments to carry a heavier burden than otherwise in correcting sub-union imbalances. It not only makes macroeconomic and exchange rate adjustment and coordination more difficult, but also shifts a portion of the adjustment burden to foreign countries.

A lack of sufficient microeconomic flexibility, structural rigidities, and national trade policies, along with macroeconomic policies, are the contributing factors to global economic imbalances that have emerged during the first few years of the new millennium (Obstfeld & Rogoff 2005). Specifically, these include structural fiscal deficits and low saving rates in the United States, a stubbornly inflexible structure in the European economy, persistent export-led growth strategy in Japan and other Asian countries, Chinese currency's peg or near peg to the dollar, and Asian central banks' willingness to hold dollar denominated bonds. A clear symptom of such imbalances is the U.S. current account deficits of over 6% of its GDP, a historical record. The total U.S. current account deficits account for 75% of the world current account surplus. A quick adjustment to balance the current account would imply a drastic change in the world economic landscape with a sharp depreciation of the dollar, a corresponding rise of the euro and the yen, and a significant jump in inflation and interest rates in the United States. The resultant contraction in the U.S. economy and slowdown in the European and Japanese economies would send shock waves through the world economic and financial system. Only

through policy changes and coordination in both macroeconomic and microeconomic arenas and in all countries involved, can the burden of adjustment be spread among all countries involved. Thus, meaningful global policy coordination relies on simultaneous coordination in all important policy areas including macroeconomic, microeconomic, and exchange rates.

### **Political Economy of Policy Coordination**

Is the desire for policy coordination consistent with the free market principle? As Frankel *et al* (1990) put it, if in the domestic economy the working of the invisible hand under competition translates independent decision-making into a social optimum, why should not the same principle apply to policy decisions by countries in the world economy? After all, if liberals promote international economic relationships that are based on market principles and individual freedom, how at the same time, can they also emphasize the positive-sum game nature of international cooperation in which all participants gain (Cohn 2000)? Some economists are indeed opposed to policy coordination based on a belief in a decentralized, market oriented international monetary system (Corden 1981; Vaubel 1981). However, many academic economists support the need for policy coordination at the level of national governments for the mutual economic gain for all countries involved. This faith may not be in conflict with their professed faith in the liberalized market economy and decentralized decision-making after all. The reason lies in the fact that the world as we know it does not conform as an ideal place for a market economy since nation-states function as a hindrance to perfect mobility of resource allocation and maximization of world economic efficiency. The most that the

liberals can hope for is for liberal-minded governments to reduce the barriers to free exchange across countries. Given this, international economic coordination is essentially a process where sovereignty is traded off for the benefits of the international division of labor (Greenspan 1988). To the liberal, international policy coordination helps break down state barriers and thus embodies the spirit of the free market.

Armijo (2001) makes distinctions among several groups of liberal economists. The *laissez-faire liberalizers* are conservative intellectuals and managers of private financial institutions who argue for minimum government interference and regulation in international financial and monetary affairs. They advocate for the abolishment of any formal international financial management institutions such as the World Bank and International Monetary Fund. Given this, it would be impossible for this group of radical conservatives to suggest any formal coordination among national governments. However, this group is split between those advocating a gold standard or gold exchange standard and those who prefer flexible exchange rates. The two stands appear different, but both originate from their strong opposition to the discretionary actions of monetary authorities.

The *transparency advocates* are members of various official study commissions, intellectual forums, study groups, and mainstream policy analysts and academics. They emphasize the role of structural problems in causing financial imbalance or even crisis, especially in developing economies. For them, fixed exchange rates, especially overvalued exchange rates, are always part of the structural problems that have led to financial instability in

developing countries. By implication, flexible exchange rates will help provide solutions to these countries' problems. However, this group does not shun government's role in financial and monetary affairs. They advocate regulation and financial transparency (thus the name), and, by extension, certain forms of policy coordination.

The third group of liberals, the *financial stabilizers*, consists of prominent Keynesian economists, deserters of the transparency advocates, and national governments outside OECD countries, especially developing countries. While the transparency advocates focus on national economic analysis, financial stabilizers look at the international economy and advocate active exchange rate collaboration, regional monetary integration, global regulation, and even supranational organization (such as a global lender of last resort).

Although the latter two groups of liberals hold views on policy coordination consistent with their liberal economic perspective, the practice in international policy coordination seems to favor the perspective of the realist. Realists believe that in international relations, including economic relations, the state seeks to strengthen the power of the nation, even at the expense of other countries. The state maneuvers for relative gain over other nations. International cooperation, including economic policy coordination, is seen as a zero-sum game that can occur only when powerful states are willing participants for their own benefit. In practice, policy coordination, as some point out, seems to confirm the view of the realists in that a hegemonic power is the precondition for a stable international economic order (Cohn 2000; Ingham 1994).

For example, the participation by the United States, in support of its own political

and economic agenda, is regarded as the crucial reason for international coordination during the early post WWII period, as reflected in the Bretton Woods regime. Since the 1990s the U.S. has acted as the sole super power and has had a political agenda different from that during the Cold War era. The result has been that the U.S. need for coordination has diminished, contributing to the declining frequency of policy coordination. This contradicts the ground rule for policy coordination actions that all participants benefit, which would ultimately limit the maneuvering room for powerful states intent on their own gain at the expense of others.

Even within liberal-minded economists and policymakers, there are diverging views between the Keynesians and monetarists, and between fiscal demand management and independent central banking. Canzoneri *et al* (2005) characterize the models used in policy coordination studies as Keynesian in nature. This is essentially because the widely adopted macroeconomic transmission mechanism works in a way that a policy-induced or autonomous income expansion in one country will eventually translate into income expansion in other countries. In contrast, the Chicago school monetarists argue against exchange rate and monetary coordination (Canzoneri *et al* 2005), believing that flexible exchange rates will insulate domestic income and employment from foreign economic disturbances, including foreign monetary policy.

Although economic reality since the end of Bretton Woods has proved otherwise, the monetarist perspective has become dominant in macroeconomic policymaking. A particular manifestation is the changing role of finance ministers and central bankers in macroeconomic policymaking. Elgie and

Thompson (1998) identify two major periods of central banking independence.

The first period, from 1914 to early 1970s, was characterized by central banking decision-making that was largely taken over by the government. During this period, there was increasingly greater government involvement in the economy, in which monetary policy played a significant role in demand management by accommodating expansionary fiscal policies. Polillo and Guillen (2005) hold that finance ministers were the key decision makers under Bretton Woods while central bankers played a relatively limited and quiet role in economics and finance policymaking, operating as yet another state agency.

The second period of central banking independence, from the collapse of Bretton Woods onward, has seen the increasing independence of central banking decision-making (Elgie and Thompson 1998). During this period, a strongly held consensus has emerged (Goodman 1992) that there is no long term tradeoff between employment and inflation. Any long run monetary expansion will lead to inflation. As a result, fiscal policy has taken a back seat in macroeconomic policymaking while monetary policy with price stability as a mandate has taken precedent (Bergsten and Henning 1996). The trend toward independent central banking reflects the attempt by government to engineer conditions of price stability without taking any blame for short-run restrictive policies which may be required (Buiter 2004; Baines 2001).

A general consensus among the G-7 countries is that the optimal policy making is to pursue sustainable policies at home and thus refrain from direct exchange rate management. This corresponds to a paradigmatic shift in macroeconomic



policymaking, from Keynesianism to monetarist neoliberalism (Polillo & Guillen 2005). Polillo and Guillen (2005) argue that after the collapse of the Bretton Woods agreement, the industrialized countries tried a variety of approaches toward stabilization in the world economy, including the G-7 process and policy coordination, which Siebert labels as the Keynesian coordinated demand management (Bergsten & Siebert 2003). Toward the late 1980s and early 1990s, neoliberalism in macroeconomic policymaking became the dominant thinking, which gave independent central banking and pursuing sustainable domestic policy priority. This may well explain the substitution of independent monetary policymaking for explicit policy coordination since the 1990s.

Some point out the changing world economy as the foundation for the shift in macroeconomic policymaking and therefore the pattern of policy coordination (Baines 2001; Webb 1995). In particular, many have argued that the increasing capital mobility in the world economy is a destabilizing factor in that it hinders the effort of policy coordination among national governments. The so-called “unholy trinity” illustrates this point clearly (Baines 2001; Cohen 1998). According to Mundell (1963) and Fleming (1962), the three key policy objectives of government, exchange rate stability, capital mobility, and domestic macroeconomic policy autonomy can not be simultaneously achieved.

Only two of these three desiderata can be achievable at any time. For example, when capital mobility is high and a country pegs its exchange rate to another country's currency, its domestic interest rates will be linked to foreign interest rates, which severely limit its ability to pursue an independent domestic monetary policy. A

tightening of domestic monetary policy that raises domestic rates above foreign rates also induces capital inflows in response to the cross-border return differential; this dampens the initial rise in domestic interest rates and induces lower domestic demand for imported goods, which further dampens the contractionary effects of the higher interest rate. Thus, under a pegged exchange rate, control of domestic interest rates and the aggregate demand impact of monetary policy are limited. In addition, in defending the peg, a country may be forced to accumulate reserves and suffer from inflation, or risk reserve and capital flights and suffer from high interest rates (Parry 1998).

Essentially, capital mobility in the world economy has made fixed exchange rate systems increasingly difficult to manage when policies are largely dictated by their own domestic agenda. Many believe that this is the crucial factor that brought down the Bretton Woods System (Baines 2001; James 1996). The high degree of capital mobility in today's world economy has constrained government freedom in pursuing domestic policy and caused international economic instability (Webb 1995). Government's efforts in pursuing intermediate exchange rates are likely to prove futile (Baines 2001).

Even within a monetary union, tensions can arise between the monetary authority and the fiscal authorities. For example, the EMU delegate monetary policy to a single authority, ECB while the national executive authorities are left with the power to design their own fiscal policy, though the fiscal authorities' power was curtailed by the Stability and Growth Pack. The fiscal power of national governments affects price stability, financial stability, or the EMU's external balance directly or indirectly

through the ECB's reaction to national economic policies. Such a structure thus creates new and amplifies existing externalities of economic policies among member states. It also creates free-riders in that national governments have less incentive to consider the consequences of their national policy for price stability and external balance (Hagen & Mundscheink 2003). There have been constant conflicts between the monetary authority and the national fiscal authorities. Each side lectures the other on the performance of their duties. The ECB admonishes finance ministers on the need for tighter fiscal policy and structural reforms instead of hiding behind a common currency and hoping for a free ride. This illustrates the possibility that international coordination may be used as a device to divert attention away from necessary domestic reforms to correct economic imbalance.

On the other hand, finance ministers and politicians in the euro zone remind the ECB to boost employment and income through lowering interest rates (Buiter 2004; Aarle *et al.* 2002). Further theoretical study confirms the difficulties in reaching a unionwide coordination. Aarle *et al.* (2004) find that although fiscal cooperation among fiscal authorities can improve the welfare of a monetary union, a grand coordination involving monetary and fiscal authorities can only result in stable improvement when member countries' economies are symmetric. In asymmetric cases, such coordination cannot result in welfare improvement. This conclusion implies a stability problem for a monetary union with a design like the EMU since fiscal authorities and the central bank, with their different mandates, all have incentives to deviate from cooperation. Buiter (2004) points out that except in a financial crisis

and in combating deflation, effective overall coordination between the ECB and member countries' fiscal authorities is unlikely unless there is a single unionwide fiscal authority.

## Conclusion

In a world economy that is characterized by increasing integration, it is understable that the issue of international economic coordination has attracted attention from academics and policymakers. However, although there are strong theoretical arguments for the case of policy coordination, the only consensus regarding policy coordination so far is that there is a need for information sharing among national policymakers. Such a "strong argument and weak practice" situation is largely the result of the prevailing policymaking process that emphasizes the role of independent central banking and the pursuit of sustainable domestic policy as the most important approach toward macroeconomic policymaking (Dodge 2004).

The right economic policies, according to Dodge, are those that encourage sustainable growth and financial stability in domestic economies. Specifically, the monetary policy should be directed at keeping inflation low and stable and fiscal policy should be disciplined in terms of public finance. In addition, policies should foster structural changes in favor of flexibility and should encourage trade liberation. With such a set of policies, in conjunction with dialogue and exchange of information among policymakers, the scope of policy coordination is limited in that the economies will take care of themselves.

The new open macroeconomic models suggest that gains of policy coordination are limited and only a second-order problem compared with the overall gains from

macroeconomic stabilization (Obstfeld and Rogoff 2002). This theoretical conclusion supports the G-7 central bankers' stand that the priority in macroeconomic policymaking is the pursuit of sustainable domestic policy. As Baine (2001) points out, macroeconomic policymaking is affected by perspective as well as actual macroeconomic conditions.

Given the prevailing monetarist perspective in macroeconomic policymaking and the central banking independence, in conjunction with a world economy free of major financial crisis since the late 1990s, major actions in macroeconomic policy coordination will largely be confined within the dialogue and exchange of information in the near future. Policy coordination as a result of monetary integration will likely be hampered by a lack of sufficient microeconomic flexibility and/or symmetries.

### **Selected References**

- Aarle, B. van, G.D. Bartoleomeo, J. Engwerda and J. Plasmans. (2002) "Monetary and Fiscal Policy Design in the EMU: an Overview", *Open Economies Review*, Volume 14, pp. 321-340.
- Aarle, B. van, G.D. Bartoleomeo, J. Engwerda and J. Plasmans. (2004) "Policymaker's Coalition and Stabilization Policies in the EMU", *Journal of Economics*, Volume 82, pp 1-24.
- Anonymous. (2005) "Speeding up European Reforms: A Master Plan for the Lisbon Process", *CESifo Forum*, Volume 6, pp. 21-45.
- Armijo, L.E. (2001) "The Political Geography of World Financial Reform: Who wants What and Why?", *Global Governance*, Volume 7, pp. 379-396.
- Baines, A.C. (2001) "Capital Mobility, Perspectives, and Central Bank Independence: Exchange Rate Policy since 1945", *Policy Sciences*, Volume 34, pp 171-193.
- Barrell, R., K. Dury and I. Hurst. (2000) "International Monetary Policy Coordination: An Evaluation Cooperative Strategies using a Large Econometric Model", *Working Paper* 160. London: National Institute of Economic and Social Research.
- Bergsten, C.F. and H. Siebert. (2003) "Should G-7 Policy Coordination be Revived?", *The International Economy*, Volume 14, pp Volume. 18-23.
- Bloomfield, A.I. (1959) *Monetary Policy under the International Gold Standard: 1880-1914*. New York: Federal Reserve Bank of New York.
- Bofinger, P. and T. Wollmershauser. (2003) "Managed Floating as a Monetary Policy Strategy", *Economics of Planning*, Volume 36, pp. 81-109.
- Branson, W. and C. Healy. (2005) "Monetary and Exchange Rate Policy Coordination in ASEAN +1", *National Bureau of Economic Research Working Paper Series* 11713.
- Bryant, R.C. (1995) *International Coordination of National Stabilization Policies*. Washington D.C.: Brookings Institution.
- Buiter, W.H. (2004) "Two Naked Emperors? Concerns about Stability & Growth Pack and Second Thoughts about Central Bank Independence", *Fiscal Studies*, Volume 25, pp 249-277.
- Canzoneri, M.B., R.E. Cumby and B.T. Diba. (2005) "The Need for International Policy Coordination: What's Old, What's new, What's Yet to Come?", *Journal of International Economics*, Volume 66, pp. 363-384.
- Caporale, G.M., M. Chui, S.G. Hall and B. Henry. (2003) "Evaluating the Gains to

- Cooperation in the G-3", *Empirica*, Volume 30, pp. 337-356.
- Caves, R.E., J.A. Frankel and R.W. Jones. (1999) *World Trade and Payments: An Introduction*. Eight Edition. Reading, MA: Addison-Wesley.
- Chaplygin, V. (2004) "Russian and Belarus Monetary Union: Phantom or Reality?", *Intereconomics*, Volume 39, pp. 29-35.
- Coeure, B. and J. Pisani-Ferry. (1999) "The Case against Benign Neglect of Exchange Rate Stability", *Finance & Development*, Volume 36, pp. 5-8.
- Cohen, B.J. (1998) *The Geography of Money*. Ithaca and London: Cornell University Press.
- Cohn, T.H. (2000) *Global Political Economy: Theory and Practice*. New York: Longman.
- Corden, W.M. (1981) "The Logic of International Monetary Non-system", in F. Machlup, G. Fels, and H. Muller-Groeling (Editors), *Reflections on a Troubled World Economy: Essays in Honor of Herbert Giersch*. London: St. Martins Press.
- Corsetti, G. and P. Pesenti. (2001) *International Dimensions of Optimal Monetary Policy*. Working Paper 8230. Cambridge, Mass: National Bureau of Economic Research.
- Diaz-Roldan, C. (2004) "International Monetary Policy Coordination under Asymmetric Shocks", *International Advances in Economics Research*, Volume 10, pp. 72-82.
- De Grauwe, P. and F.P. Mongelli. (2005) "Endogeneities of Optimal Currency Area: What Brings Countries Sharing a Single Currency Closer Together?", *Working Paper Series No. 468*. Frankfurt: European Central Bank.
- Dodge, D. (2004) "Co-operation and the Conduct of Economic Policy", *Bank of Canada Review*, Summer, pp. 57-61.
- Eichengreen B. and O. Jeanne. (2000) "Currency Crisis and Unemployment: Sterling in 1931", in P. Krugman (Editor), *Currency Crisis*. Chicago: The University of Chicago Press.
- Fasano, U. and Z. Iqbal. (2002) Common Currency, *Finance & Development*, Volume 39, pp. 42-45.
- Fernandez-Arias, E. and P. Montiel. (2001) "Reform and Growth in Latin America: All Pain, No Gain?", *International Monetary Fund Staff Papers*, Volume 48.
- Ferrando, G.M.L. (2000) *Monetary Policy Coordination, Monetary Integration and Other Essays*.  
<http://levine.sscnet.ucla.edu/archive/licandrothesis.pdf>
- Financial Times. (2006a) "Moscow under Pressure on Gas Supplies" *Financial Times*, February 10. London: Financial Times Ltd.
- Financial Times. (2006b) "Asia Trade a "Boodle Bow" of Competing Pacts, Says ADB Chief", *Financial Times*, February 9. London: Financial Times Ltd.
- Fischer, S. (1988) International Macroeconomic Policy Coordination", in M. Feldstein (Editor), *International Economic Cooperation*. Chicago: the University of Chicago Press.
- Fleming, J.M. (1962) "Domestic Financial Policies under Fixed and under Flexible Exchange Rates", *IMF Staff Papers*, Volume 9, pp. 369-379.
- Frankel, J., S. Erwin and K. Rockett. (1992) "A Note on Internationally Coordinated Policy Packages Intended to be Robust under Model Uncertainty", *American Economic Review*, Volume 82, Number 4, pp 1052-56.
- Frankel, J.A., M. Goldstein and P.R. Masson. (1990) "The Rationale for, and Effects of, International Economic Policy Coordination", in W.H. Branson, J.A. Frankel, and M. Goldstein (Editor), *International Policy Coordination and*

- Exchange Rate Fluctuations*. Chicago: Chicago University Press.
- Frankel, J.A. (1988) "The Obstacles to Macroeconomic Policy Coordination with an Analysis of International Nominal Targeting (INT)", *Studies in International Finance*, Number 64, Princeton University.
- Gagnon J.E. (2004) "Comment on Edwin M. Truman", in H. Siebert (Editor), *Macroeconomic Policies in the World Economy*. New York: Springer-Verlag.
- Gutierrez, M.J. (2003) "Monetary Cooperation among Countries", in O. Bajo-Rubio (Editor), *Macroeconomic Policy in an Open Economy: Application of the Mundell-Fleming Model*. New York: Nova Science Publishers.
- Greenspan, A. (1988) "Prospect for International Economic Cooperation", in M. Feldstein (Editor), *International Economic Cooperation*. Chicago: the University of Chicago Press.
- Hagen, J. von and S. Mundschenk. (2001) "The Functioning of Economic Policy Coordination", *Working Paper B-08*. Centre for European Integration.
- Hagen, J. von and S. Mundschenk. (2003) "Fiscal and Monetary Policy Coordination in EMU", *International Journal of Finance and Economics*, Volume 8, pp. 279-295.
- Hamada, K. (1976) "A Strategic Analysis of Monetary Interdependence", *Journal of Political Economy*, Volume 84, pp. 877-895.
- Ishii, N., W. McKibbin and J. Sachs. (1985) "The Economic Policy Mix, Policy Cooperation, and Protectionism: Some Aspects of Macroeconomic Interdependence among the United States, Japan, and Other OECD Countries", *Journal of Political Modeling*, Volume 7, pp. 533-572.
- Kawai, M. and S. Takagi. (2005) "Toward a Regional Monetary Cooperation in East Asia: Lessons from Other Parts of the World", *International Journal of Finance and Economics*, Volume 10, pp. 97-116.
- Ingham, J. (1994) "States and Markets in the Production of World Money: Sterling and the Dollar", in S. Corbridge, N. Thrift, and R. Martin (Editors), *Money, Power, and Space*. Oxford, UK: Blackwell.
- James, H. (1996) *International Monetary Cooperation since Bretton Woods*. Washington D.C.: International Monetary Fund.
- Krugman, P.R. and M. Obstfeld. (1997) *International Economics*. Fourth Edition. Reading, MA: Addison-Wesley.
- Lowell, L. (1994) *Why (do) We Want Macroeconomic Policy Coordination?* Santa Monica, CA: Rand.
- Marston, R.C. (1988) "Exchange Rate Policy Reconsidered", in M. Feldstein (Editor), *International Economic Cooperation*. Chicago: the University of Chicago Press.
- Masson, P. and A. Pattilo. (2004) "A Single Currency for Africa?", *Finance & Development*, Volume 41, pp. 8-15.
- McKibbin, W. (1997) "Empirical Evidence on International Macroeconomic Policy Coordination", in M.U. Fratianni, D. Salvatore, and J. von Hagen (Editors), *Handbook of Comparative Economic Policies, Volume 5: Macroeconomic Policies in Open Economies*. Westport, CT: Greenwood Press.
- Meyer, L.H., B.M. Doyle, J.E. Gagnon and D.W. Henderson. (2002) "International Coordination of Macroeconomic Policies: Still Alive in the New Millennium?", *International Finance Papers*, Number 723. Washington, D.C. Board of Governors of the Federal Reserve System.

- Mundell, R.A. (1961) "A Theory of Optimum Currency Areas", *American Economic Review*, , Volume 51, pp. 657-65.
- Mundell, R.A. (1963) "Capital Mobility and Stability Policy under Fixed and Flexible Exchange Rates", *Canadian Journal of Economics*, , Volume 29, pp. 475-485.
- Mussa, M. (1979) "Macroeconomic Interdependence and the Exchange Rate Regime", in R. Dornbusch and J. Frankel (Editors), *International Economic Policy*. Baltimore: Johns Hopkins University Press.
- Obstfeld, M. and K. Rogoff. (2000) "New Directions for Stochastic Open Economic Models", *Journal of International Economics*, , Volume 50, pp. 117-153.
- Obstfeld, M and K. Rogoff. (2002) "Global Implications of Self-oriented National Monetary Rules", *Quarterly Journal of Economics*, , Volume 117, pp. 503-536.
- Obstfeld, M and K. Rogoff. (2005) "Global Current Account Imbalances and Exchange Rates Adjustments", *Brookings Papers on Economic Activity*, pp. 1-77.
- Oudiz, G. and J. Sachs. (1984) "Macroeconomic Policy Coordination among the Industrial Economies", *Brookings Papers on Economic Activity*, Number 1, pp. 1-64.
- Parry, Robert. (1998) *Conducting a Monetary Policy in a Global Economy*. [www.findarticles.com/p/articles/mi\\_m1094/is\\_n4\\_v33/ai\\_21222055](http://www.findarticles.com/p/articles/mi_m1094/is_n4_v33/ai_21222055)
- Polillo, S. and M. Guillen. (2005) "Globalization Pressures and the States: the Worldwide Spread of Central Bank Independence", *The American Journal of Sociology*, Volume 110, pp. 1764-1802.
- Rogoff, K. (2003) "Globalization and Global Disinflation", *Economic Review*, (Federal Reserve Bank of Kansas City) Volume 88, pp. 445-478.
- Sachs, J. and W. McKibbin. (1985) "Macroeconomic Policies in the OECD and LDC External Adjustment", *Discussion Paper 25*. Washington DC: Brookings Institution.
- Schultze, C.L. (1988) "International Macroeconomic Policy Coordination-Marrying the Economic Models with Political Reality", in M. Feldstein (Editor), *International Economic Cooperation*. Chicago: the University of Chicago Press.
- Sutherland, A. (2002) "International Monetary Policy Coordination and Financial Market Integration", *Board of Governor of Federal Reserve System International Finance Working Papers* No. 751.
- Temprano-Arroyo, H. (2003) "Prospect for Regional Monetary Integration in Latin America: a View from the EU", *Comparative Economic Studies*, Volume 45, pp. 348-420.
- Truman, E.M. (2004) "A Critical Review of Coordination Effort in the Past", in H. Siebert (Editor), *Macroeconomic Policies in the World Economy*. New York: Springer-Verlag.
- Vaubel, R. (1981) "Coordination or Competition among National Economic Policies?", in F. Machlup, G. Fels, and H. Muller-Groeling (Editors), *Reflections on a Troubled World Economy: Essays in Honor of Herbert Giersch*. London: St. Martins Press.
- Webb, M. (1995) *The Political Economy of Policy Coordination: International Adjustment Since 1945*. Ithaca and London: Cornell University Press.

Bin Zhou

Department of Geography  
Southern Illinois University  
Edwardsville, Illinois, USA  
bzhou@siue.edu



## Government Debt

*Matias Vernengo*

### **Introduction**

Most economists in recent decades would agree that future generations bear a burden resulting from public spending financed internally by public debt which takes place in the current period. The burden of public expenditure financed by debt is the reduced private consumption supposedly required by the withdrawal of resources from the private sector. The view of government debt as a burden on society, particularly an intergenerational burden, is relatively new, however, and derives fundamentally from the work of James Buchanan starting in the 1950s and consolidated around the 1970s in the midst of the crisis of Keynesian economics.

In fact, in historical perspective economists' views on whether debt is a blessing or a burden have changed significantly. During the eighteenth and nineteenth centuries dominated by the principles of sound finance, and balanced budgets, government debt, albeit in moderation, was seen as blessing by most layman and politicians. For example, Alexander Hamilton the first Treasury Secretary and main strategist of the economic system in the post-independent United States did in fact argue for the development of a national public debt as a form of providing a strong central government. Hamilton saw government debt and a national public bank as the pillars of the British economic success that he wanted to emulate in America. Economists, on the other hand, saw public debt with apprehension.

This changed in the early part of the twentieth century. The Keynesian

Revolution, and even more emphatically Abba Lerner's functional finance theory, provided the rationale for the idea that public debt would be good in a crisis. Functional finance established that, under certain conditions, government debt is an essential feature of a healthy economy. Within the mainstream of the profession, it was also argued that debt was a useful policy instrument. In a general equilibrium system, with incomplete markets, public debt may assist capital formation and lead to higher rates of growth. Politicians, on the other hand, were not quite convinced about the positive effects of debt.

However, starting in the 1950s, the benign view of public debt gave gradually away to a more pessimistic position, partly as a result of the growth in non-war related debt. Also, the process of globalization—the liberalization of movements of goods, services and capital—since the end of Bretton Woods, led to the proliferation of tax heavens, and an increase in tax competition making public debt more costly to service. Finally, the stagflation of the 1970s, and the conservative revolution that followed fed a tax revolt that made increasing tax burdens to finance public debt unlikely. Measures to contain spending, and to reduce the burden of debt became more common. The most well known policy measure to curb debt is undoubtedly the Maastricht limits on public debt to Gross Domestic Product (GDP) ratios imposed on the European countries.

This entry will discuss the historical evolution of public debt, and the theoretical discussions regarding the effects of government debt on economic activity. It should be noted that this entry is only concerned with domestic public debt. It is generally accepted that external debt is a burden on the community, forcing a real

transfer of goods and services from the debtor to the creditor countries.

### **From the Tax State to the Debt State**

The rise of the tax State and the rise of modern capitalism are part of the same process (Schumpeter 1918; Musgrave 1980-81). The expansion of markets and the development of long distant trade created new sources of revenue from the thirteenth century on. Indirect taxes such as customs and excise taxes were initially the main source of revenue. Only in the late nineteenth century and more clearly in the twentieth century did direct taxes on income and to a lesser extent on wealth become relevant. Tax revenues remained relatively low, averaging ten per cent by the late nineteenth century in the main developed countries (Tanzi & Schuknecht 2000:52-53).

In normal conditions the ability to spend was directly restricted by tax revenues. The development of the financial system meant that States could borrow and spend beyond tax revenues. The only limitation would be that banks should be willing to lend, that is the State had to be creditworthy. The Italian City States—in particular Florence, Genoa and Venice—were the first to develop sophisticated markets for public debt. These cities were the first to overcome the power of their medieval masters – the surrounding landowners, the pope and the more distant but still powerful Holy Roman Emperor. Also, their geographical position was strategic to dominate the trade between the northern European cities and the East. Several luxury commodities come from the East (e.g. silk, spices, rugs, etc) through the silk and spice trade routes. By dominating Mediterranean trade the Italian City State became quite wealthy.

Public borrowing, and the development of secondary markets for public loans, meant

that the ability to spend – for the most part associated with military expenses at that point – was greatly increased. Repayment of public loans was limited to the ability to raise taxes from its own population or, in cases of war, to the capacity to extract booty from defeated enemies. The most important forms of taxes on the domestic population were excise (e.g. a tax on salt) and custom taxes (on imports and exports). Also one must not forget that interest rates had to be paid on outstanding loans. Interest rates were relatively low reaching four per cent on some issues (Homer & Sylla 1996:104-132).

The power of the State to tax and to borrow in financial markets should not be exaggerated. Financial markets maintained a fair amount of autonomy. If we look at the long period of development in the Western world that starts with the Italian City States, and goes through successive cycles of hegemonic power—including the Portuguese and Spanish empires, the Dutch economy, the British empire, and culminating with American hegemony—we find out that it has been only during a limited amount of time that individual States imposed their sovereignty over money and the international monetary system (Rochon & Vernengo 2003).

The commercial and the subsequent industrial revolutions went hand in hand with the development of national States. The unprecedented rates of growth in Western Europe from the late eighteenth century onwards—in particular after the Industrial Revolution—are associated to the alliance of the merchant bourgeoisie and the nascent nation State. The alliances were certainly not homogenous, and Fernand Braudel distinguished between commercially oriented City States and national territorial States to highlight the differences. Some States owed their existence to trade others



simply despised commercial activities. All successful States were able to control a booming domestic market.

In its early stages, public debt should be seen as a form of repayable taxation – that is, instead of taxing the wealthy, which would mean that they have no claim over the money paid to the State, the well to do were forced to lend money to the State, in which case they had a claim on their money at some point in the future. This was a loophole on the tradition of not taxing the wealthy. Homer and Sylla (1996:74) argue that the first State loans, after nine centuries of dark ages for financial markets following the fall of the Roman Empire, levied on the wealthy citizens of Florence, Genoa and Venice “were a form of tax [where] the taxpayer, however, got a claim to interest as a return on his payment.” The notion of repayable taxes was in fact introduced to protect those that were originally free from taxes, i.e. the landed aristocracy (MacDonald 2003:84).

The credit record of the Italian City States—and of almost all the other hegemonic powers that followed—was quite good. However, defaults did occur, and creditors tended to be more confident in the ability and the willingness to repay whenever a system of checks was in place. A bureaucratic system of tax collection and a transparent process of parliamentary budget making were then essential institutional mechanisms, perfected by the British that allowed the expansion of public credit to new levels (Ferguson 2001:112). In fact, at the end of the Napoleonic Wars the debt to GDP ratio in the UK was at almost three hundred per cent (p. 126). That level of debt accumulation is impressive even for modern standards.

However, it is clear that a price had to be paid in order to increase the ability of the

State to tax and borrow, and hence increase its capability to spend larger sums. In general higher levels of public borrowing were only possible when some kind of parliamentary representation was in place, hence reducing the absolute power of the prince. Hence the power of purse came at the expense of political power, and one may claim that public debt is one of the factors leading to more democratic societies (MacDonald 2003). Also, ability to raise the public debt to higher levels at lower cost in part explains international hegemonic power, as the final victory of Britain over France in the Napoleonic Wars demonstrates (Ferguson 2001:174).

International hegemonic power and a more democratic society, in which the prince shares power with his subjects, implies that the State must be able to keep pace with the expenses of war and with the rising social demands of its population. In fact, in Britain the process of industrialization and urbanization, with the rise of an urban working class, and the rise to dominance in international markets, leads to a whole new set of military and social expenses, that could only be satisfied by increasing taxation – if the stock of debt was to be reduced. Slowly throughout the nineteenth century the Night Watchman State – that is, the State that sits and does nothing – evolved into a more active participant in the economy. By the late nineteenth-century—with the rise of socialism and the labor movement in Western Europe—social concerns become an essential part of policy preoccupations. Starting with Bismarck in Germany a limited system of welfare was developed.

Balanced budgets were seen as essential, not just to keep the debts at sustainable levels, but also to avoid excessive burdens on the poor. William Gladstone—the prime

minister that dominated British politics in the second half of the nineteenth-century—believed that the miseries of the poor stemmed from the unnecessary burden of indirect taxes, which are paid mostly by the poor (Maloney 1998:32). For this very reason he fought for the imposition of income taxes.

With the depression and the advent of Keynesian economics the role of the State changed significantly. Income tax and social security contributions become the main source of revenue, and the role of indirect taxes diminished. A progressive income tax—once advocated by Marx and Engels (1948:28) as a radical measure—became standard in the Western world. The tax State had gone full circle, from its modest origins when public borrowing was a substitute for taxing the privileged, and only indirect taxes, essentially on the poor were the main source of revenue, to a progressive system of taxation where the bulk of contributions came from direct taxes on the upper class. In sum, the rise of the tax State can be seen as a struggle over who would carry the burden of taxation; a battle, one should add, that was progressively, if slowly, won by the underdogs.

By the late 1970s tax revenues in the developed countries ranged from slightly less than thirty to more than sixty per cent of GDP, and some authors foresaw the downfall of the tax State. According to O'Connor (1973) a structural gap developed at the heart of the tax State. The ability to raise taxes was limited, and the social demands were virtually unlimited, hence a fiscal crisis was inevitable. The tax State as capitalism itself for Marxists authors—carries the seeds of its own destruction. Tanzi and Schuknecht (2000)—although falling short of claiming the imminent demise of the tax State—argue that the

increase in the size of government, and the pervasive fiscal deficits in the last fifty years did not lead to any social or economic advantages. According to them, since the 1960s the size of the State increased without leading to higher levels of income per capita or a better distribution of income. As a result in their view, the tax burden and expenditures should be considerably scaled down. The Tax State has given place to the Debt State (Green 1993).

The impressive increase in the flows of trade and capital that came to be denominated globalization, are seen by many authors as having dealt the last blow to the Welfare State, and leading eventually to the demise of the tax State. In particular, tax competition seems to erode the tax base, while globalization requires a more comprehensive welfare safety net. Globalization then forces higher spending while it erodes the ability to obtain higher revenues, leading to a fiscal crisis. Also, the previously fiscally friendly demographic conditions, with several contributors and few recipients of pension benefits will change in the next quarter of a century in most developed countries, increasing the burden of payments, and making a fiscal crisis more likely.

But, like the proverbial decease of Mark Twain, rumors of the death of the tax State may be exaggerated so far. In fact, there is no clear evidence of the death—a severe reduction of the size—of the tax State, at least in macroeconomic terms. The last thirty years, usually associated with globalization, have seen an increase in the burden of debt.

Regarding the size of the debt burden, Table 1 shows that from 1870 to the end of the 20<sup>th</sup> century debt to GDP ratios increased up to World War II. Further, one can see that debt to GDP ratios decreased in the post war

period until the 1970s and started increasing afterwards. The data is restricted to the G-7,

but is representative of the evolution of debt burdens around the globe (Masson and Mussa, 1995).

*Table 1. Burden of Debt (% of GDP)*

	1870	1913	1920	1937	1970	1980	1990	1997
Canada	--	--	40.6	74.1	51.9	44	72.5	93.8
France	51.4	66.5	136.8	137.2	53.1	30.9	40.2	64.6
Germany	--	--	4.2	17.4	18.4	31.1	45.5	65
Italy	92	70.6	91.3	95.7	41.7	58.7	104.5	121.7
Japan	0.6	53.6	25.6	57	12.1	51.2	65.1	87.1
UK	40.2	30.4	132	188.1	81.8	54	39.3	60.3
US	43.9	2.5	31	43.7	45.4	37	55.5	61.5

*Source:* Adapted from Tanzi & Schuknecht (2000)

High levels of public debt, some fear, may reduce—for political reasons mainly—the ability of States to run deficits in times of crisis. The impressive growth of the burden of debt in the last two centuries in most developed countries went hand in hand with the process of industrialization and capital accumulation that raised the welfare levels of their populations. However, the increase in public debt led to rising fears of a collapse of the Tax State. The question of whether government debt is a blessing or a burden on society is a crucial one that economists have been groping with for the last two centuries at least, as we will discuss in the following section.

### **Burden or Blessing?**

Economists' views on the effects of government debt have swung back and forth over the years. This was also true in the political arena. For example, with the foundation of the National Debt in 1693, and the Bank of England in 1694, public finance became a major topic in political economy debates in England (Hamilton 1947). As a

general rule, conservatives saw the debt as an inappropriate increase in the power of the crown and a threat to individual liberties. On the other hand, Whigs saw public debt as the basis for the development of the financial sector, and the expansion of trade. The political conflict can be seen as a dispute between creditor's interests (Whigs) versus landed debtor's interests (Tories).

Among the founding fathers of political economy, David Hume and Adam Smith expressed clear ideas about government debt. The usual identification of Hume as a Tory—in particular because of his historical writings—means that his views on public debt are usually associated to the conservative country party. Hence, Rotwein (1955:lxxxvi) in his authoritative introduction to Hume's writings on economics argues, "what he [Hume] seeks to show is that the inevitable continued rise in debt will not only have the gravest consequences for society but will ultimately terminate in total bankruptcy."

While it is clear that Hume was concerned, and was critical of the increasing

size of public debt, he was not a dogmatic critic of the fiscal policies of the crown. He clearly refers to the ill resulting from public debt, but he also claims that some good results from a national debt (Hume 1955:90-107). Adam Smith was also more concerned about the dangers of bankruptcy than the possibilities for expanded trade that government debt facilitated (Smith 1776). However, one must note that Smith thought that taxes hindered capital accumulation in a way that government debt did not. As noted by Winch (1998:3) McCulloch criticized both, Hume and Smith, for not perceiving that output growth made the burden of debt easier to carry.

David Ricardo contributions to the question of public finances have cast a long shadow. Ricardo was even more concerned than his predecessors with the negative effects of taxes and public debt on capital accumulation. He considered the British debt, together with the Corn Laws, as two of the great evils plaguing the nation (Winch 1998:8). For Ricardo (1817:150-52) taxes either reduce capital accumulation or force taxpayers to reduce consumption. Debt, as deferred taxes were argued to have the same effect. Hence, Ricardo (being a Radical connected with Jeremy Bentham and James Mill), was for reducing government debt, and maintaining a relatively low level of taxes. Such measures at the time were opposed by both Whigs and Tories. Ricardo theory was later revived as the Ricardian equivalence theorem (Barro 1974).

Thomas Malthus—personal friend and main intellectual foe of Ricardo—not surprisingly held different views regarding government debt. As a moderate Whig, Malthus was against paying down public debt, on the grounds that it allowed debt holders to expand consumption and effectual demand, a result that was prevented by Say's

Law in the Ricardian system (Ferguson 2001:129-30).

The Marginalist Revolution of the 1870s brought significant changes to the core of economic analysis—the theory of value—but left the dominant view on government debt among academic economists unchanged. It was only the Keynesian Revolution that brought a significant change in academic circles. Thirty years after Keynes' (1936) masterpiece, Tobin (1965:679) could argue that the conventional view among academic economists—but not in political circles—was that government debt is not a burden on future generations.

Keynes (1936) was opposed to deficit spending, and hence debt accumulation, as a means of stimulating consumption. However, Keynes favored government borrowing to finance capital expenditures to stimulate investment. Lerner (1943) argued that government deficits, and the consequent accumulation of debt, should be instruments for the maintenance of full employment. This view became known as the functional finance approach to deficits and debt. Economists in the functional finance tradition argue that a high debt-to-GDP ratio, at any rate as long as the borrowing is done internally in the country's own currency, does not imply any appreciable risk for default. Nor does it limit the scope for active, counter-cyclical policy measures in any relevant sense. Moreover, it is argued that wealth effects—through which public debt fuels private spending—tend to spontaneously establish a certain level to which the debt-to-GDP ratio converges in the long run. In this view, therefore, the debt problem takes care of itself in the long run, and there is little cause for concern.

Colander (1984:1574) suggests that while logically on Lerner's side, Keynes was unwilling to endorse functional finance for

political reasons. The main political constraint on functional finance would be the possibility of persistent deficits and increasing government debt during peacetime.

The main problem raised by persistent deficits is the one of debt sustainability. Domar (1944) showed that debt to GDP ratios tend to increase at explosive rates when the rate of interest exceeds the rate of growth. This can be easily understood in the following way. The rate of interest on government bonds is the rate at which treasury debt increases, while the rate of growth of the economy is a proxy of the capacity to pay for the debt, since revenues rise in a booming economy. Consequently, if the rate of interest is higher than the rate of growth, then the burden of debt is rising faster than the ability to pay, and the proportion of debt to GDP must rise (Blanchard 1990; Galbraith & Darity 1995).

A simple way to assess the burden of debt is to study the debt servicing requirement vis a vis the capacity to service the debt. A typical measure is the debt to GDP ratio. We define the debt-GDP ratio as

$$(1) \quad d = D/Y$$

where  $D$  is the inherited stock of debt, and  $Y$  is the flow of income, that is, the Gross Domestic Product. It is quite obvious that the rate of growth of  $d$  is the difference between the rates of growth of  $D$  minus the rate of growth of  $Y$ . That is

$$(2) \quad \frac{\dot{d}}{d} = \frac{\dot{D}}{D} - \frac{\dot{Y}}{Y}$$

Further, the change in the stock of debt is equal to the budget deficit plus interest payments on government debt. Formally

$$(3) \quad \dot{D} = (G - T) + iD$$

where  $G$  is government spending,  $T$  is the tax revenue and  $i$  is the rate of interest. Substituting (3) into (2), and renaming the rate of growth of income as  $g$ , we get

$$(4) \quad \frac{\dot{d}}{d} = (i - g) + \frac{(G - T)}{D}$$

According to the equation above, the rate of growth of the debt-GDP ratio is the sum of the difference of the rate of interest and the rate of growth, and the ratio of the budget deficit to the stock of debt. It is easy to see that even for a positive but constant public deficit the debt-GDP ratio will be falling in time if the rate of growth exceeds the rate of interest. There are two ways of guaranteeing debt sustainability, namely: (1) maintaining primary surpluses, that is, a negative  $G$  minus  $T$ , or (2) if the economy grows faster than the rate of interest.

Some Keynesian authors (e.g. Modigliani 1961; Musgrave 1959) argue that the method of financing government expenditures, that is, either current taxes or government debt, and hence, future taxes, may affect the uses of resources. This result, one must note, contradicts Ricardian Equivalence. In this view, tax finance displaces current consumption, while debt finance displaces current investment, and has a negative impact on capital accumulation. This consolidated the mainstream Keynesian vision that fiscal deficits and the accumulation of debt have a positive effect in the short run, generating higher levels of employment, but a long term negative effect on capital accumulation.

Note, however, that the traditional Keynesian result may be affected by market imperfections. Diamond (1965) showed that the failure of competitive markets to support

optimal intertemporal allocations can stem from two sources: (1) the divergence between the planning horizon of individuals and the horizon of economic activity, evident in economies of overlapping generations, and (2) market imperfections, such as missing markets for the allocation of risks. Gale (1990) shows that, given financial market imperfections, public debt by introducing low risk securities may assist capital formation and economic growth.

Buchanan (1958) is the most vocal mourner of Gladstonian sound finance. Buchanan blamed Keynesian economics for the elimination of the balanced budget constrained, which led to the lifting of the limits on politicians' natural bias towards budget deficits. Buchanan's ideas—which according to Tobin (1965:680) are reduced “to the assertion that payment of taxes is *per se* a burden [and,] since debt finance postpones the levy of taxes, it obviously shifts [the tax] burden to future generations”—played an important role in the conservative revolution of the late 1970s. However, most discussions about the burden of debt in academic circles are still illuminated by the old Keynesian preoccupations with the effects of debt on private incentives for investment.

### **Interest Rate Burden**

In the last three decades the Keynesian belief in the positive effects of deficits (at least in the short run) has been turned upside down. Giavazzi and Pagano (1990) argue that a fiscal contraction—reduction in government spending—could have an expansionary effect on the economy, provided that the cuts in government spending lead to a perception of permanently lower taxes. That is, if the government decides to spend less, then taxpayers would have to pay less in the future and would

increase both consumption and investment. This increase in consumption and investment may outweigh the decrease in government consumption, leading then to a growing economy and higher levels of income.

Giavazzi and Pagano (1990) put forward their argument on the basis of evidence Danish and Irish fiscal consolidation experiments in the 1980s. The recent American experience with fiscal consolidation (a reduction in deficits) during the Clinton years, and the relative prosperity that ensued, led many to believe that the Treasury and Federal Reserve bank strategy of pursuing fiscal adjustment rather than low rates of interest was behind the Clintonian expansion. This would lend support to the New Political Economy view that fiscal consolidation leads to economic growth (Drazen 2000).

Alesina et alli (1998) argue that fiscal consolidations that emphasize spending cuts rather than tax increases tend to be more successful in reduced the levels of accumulated debt. Alesina and Perotti (1995, 1997) argue using a sample of OECD countries that the most effective fiscal adjustments are those that concentrated in cutting two specific types of outlays, namely: expenditures on social programs and government wages. Ardagna (2004) argues using data from a panel of OECD countries that the success of fiscal adjustments in decreasing the debt to GDP ratio depends on the size of the fiscal contraction and less on its composition.

All the research on the effects of fiscal consolidation is associated to the idea that a lower burden of debt would lead to higher growth. However, there is very little evidence that lower debt is associated with capital accumulation. In fact, the literature is inconclusive on the question of the optimal

size of government debt. For that reason, some authors have argued that fiscal consolidation has gone too far. For example, Stiglitz (2003:49) argues that the current mood with respect to fiscal policy around the world is related to the apparent success of Clinton's fiscal consolidation, even though Clinton "pushed deficit reduction too far."

While most economists and politicians are concerned about the burden of debt—

either because it promotes unfair intergenerational transfers or because it slows down the pace of capital accumulation—post-Keynesian authors have argued about the burden of interest rates. In this view, the problem of debt accumulation results less from the increase in social spending, but from the combination of high rates of interest and low rates of growth in the last thirty years.

*Table 2. Interest Rate Burden (% of GDP)*

	1 870	1 913	1 920	1 937	1 970	1 980	1 995
Canada	--	--	4	5	1	2	4
France	.2		.8	.8	.2	.5	.3
Germany	5	--	4	5	0	1	3
Italy	.2		.5	.4	.5	.5	.2
Japan	0	0	--	0	0	1	2
UK	.3	.3		.9	.4		.5
US	4	2	4	4	1	5	1
	.5	.6	.4	.7	.8	.4	1.1
	--	--	1	--	0	2	3
			.1		.4	.4	
	1	1	5	5	2	4	3
	.7	.7	.7	.3	.7	.7	.7
	1	--	1	1	1	2	3
	.4		.3		.3	.3	.2

*Source:* Adapted from Tanzi & Schuknecht (2000)

Pasinetti (1997:168) argues that the growth of debt is not caused by Keynesian profligacy, and that the vulnerability of fiscal accounts "usually attributed to the high size reached by the debt ... is in fact due to the high level reached by interest rates." The causes of high interest rates and low rates of growth are highly complex and related to the demise of the accumulation regime that was in place during the so-called Golden Age of capitalism (Glyn et al 1990). However, Pasinetti (1997) argues convincingly that a revenge of the rentier—

the reverse of Keynes' euthanasia of the rentier—resulting from income distribution conflict is at the heart of the process.

Interest payments as a share of total GDP—which represent the burden of interest payments associated with government debt servicing—have increased, crowding out social spending (Table 2). It is the Bretton Woods period, in which capital controls were widespread, and interest rates were set with the view to generate full employment, that allowed government social spending to increase without leading to

explosive increase in government debt. That particular period vindicates the Lernerian benign view of deficits and debt.

In this view, then, the problem of a growing burden of debt should not be dealt by measures like the Maastricht limits imposed in Europe, or balanced-budget amendments, as often suggested in the US. These solutions tend to emphasize the need to curtail spending, which more often than not means social spending. A Keynesian perspective would emphasize the need for reintroducing the euthanasia of the rentier favored by Keynes, and reduce the burden of interest payments. Whether a more positive view of the effects of government debt is able to reassert itself or not will play a crucial role on the debates on public policy and will ultimately determine whether a reformed Welfare State would survive the assault of conservative forces.

### Selected References

- Alesina, A., R. Perotti and J. Tavares. (1998) "The Political Economy of Fiscal Adjustments", *Brookings Papers on Economic Activity*, Number 1, pp. 197-266.
- Alesina, A. and R. Perotti. (1995) "Fiscal Expansions and Adjustments in OECD Economies", *Economic Policy*, Volume 21, pp. 205-240.
- Alesina, A. and R. Perotti. (1997) "Fiscal Adjustments in OECD Countries: Composition and Macroeconomic Effects", *IMF Staff Papers*, Volume 44, 210-248.
- Ardagna, S. (2004) "Fiscal Stabilizations: When do they Work and Why", *European Economic Review*, Oct., Volume 48, Number 5, pp. 1047-75.
- Barro, R. (1974) "Are Government Bonds Net Wealth", *Journal of Political Economy*, Volume 82, pp. 1095-118.
- Blanchard, O. (1990) "Suggestions for a New Set of Fiscal Indicators", *OECD Working Paper*, Number 79, April.
- Buchanan, J. (1958) "Concerning Future Generations", in J. Ferguson (Editor), *Public Debt and Future Generations*. Chapel Hill: University of North Carolina Press, 1964.
- Colander, D. (1984) "Was Keynes a Keynesian or a Lernerian", *Journal of Economic Literature*, Volume XXII, pp. 1572-75.
- Diamond, P. (1965) "National Debt in a Neoclassical Growth Model", *American Economic Review*, Volume 55, pp. 1126-50.
- Domar, E. (1944) "The Burden of Debt and the National Income", *American Economic Review*, Volume 34, Number 4, pp. 798-827.
- Drazen, A. (2000) *Political Economy in Macroeconomics*. Princeton, Princeton University Press.
- Ferguson, N. (2001) *The Cash Nexus: Money and Power in the Modern World, 1700-2000*. New York: Basic Books.
- Galbraith, J. and W. Darity. (1995) "A Guide to the Deficit", *Challenge*, July-August, pp. 5-12.
- Gale, D. (1990) "The Efficient Design of Public Debt", in R. Dornbusch and M. Draghi, (Editors), *Public Debt Management: Theory and History*, Cambridge: MIT Press.
- Giavazzi, F., and M. Pagano (1990) "Can Severe Fiscal Adjustment Be Expansionary? Tales of Two Small European Countries", in O. Blanchard and S. Fisher, (Editors), *NBER Macroeconomics Annual*. Cambridge, MIT Press.
- Glyn, A., A. Hughes, A. Lipietz and A. Singh (1990) "The Rise and fall of the Golden Age", in S. Marglin and J. Schor,



- (Editors), *The Golden Age of Capitalism*. Oxford: Oxford University Press.
- Green, C. (1993) "From the Tax State to the Debt State", *Journal of Evolutionary Economics*, Volume 3, pp. 23-42.
- Hamilton, E.J. (1947) "Origin and Growth of the National Debt in Western Europe", *American Economic Review*, Volume 37, Number 2, May, pp. 118-130.
- Homer, S. and R. Sylla (1996) *A History of Interest Rates*. Third Edition. New Brunswick: Rutgers University Press.
- Hume, D. (1955) *Writings on Economics*. Madison: University of Wisconsin Press, 1970.
- Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*. New York: Harcourt Brace, 1964.
- Lerner, A. (1943) "Functional Finance and the Federal Debt", *Social Research*, Volume 10, pp. 38-51.
- Lerner, A. (1948) "The Burden of National Debt", in J. Ferguson, (Editor), *Public Debt and Future Generations*. Chapel Hill: University of North Carolina Press, 1964.
- MacDonald, J. (2003) *A Free Nation Deep in Debt: The Financial Roots of Democracy*. New York: Farrar, Straus and Giroux.
- Maloney, J. (1998) "Gladstone and Sound Victorian Finance", in *Debt and Deficits: An Historical Perspective*. Cheltenham, UK: Edward Elgar.
- Marx, K. and F. Engels. (1848) "Manifesto of the Communist Party", in L.S. Feuer (Editor), *Marx and Engels Basic Writings on Politics and Philosophy*. New York: Anchor Books, 1959.
- Masson, P. and M. Mussa. (1995) "Long Term Tendencies in Budget Deficits and Debt", in *Budget Deficits and Debt: Issues and Options*. Kansas City: Federal Reserve Bank of Kansas City.
- Modigliani, F. (1961) "Long Run Implications of Alternative Fiscal Policies and the Burden of National Debt", in J. Ferguson, (Editor), *Public Debt and Future Generations*. Chapel Hill: University of North Carolina Press, 1964.
- Musgrave, R. (1959) "Internal Debt and the Classical System", in J. Ferguson, (Editor), *Public Debt and Future Generations*. Chapel Hill, University of North Carolina Press, 1964.
- Musgrave, R. (1980-81) "Theories of Fiscal Crises: An Essay in Fiscal Sociology", in M. Boskin (Editor), *The Economics of Taxation*. Washington DC: Brookings.
- O'Connor, J. (1973) *The Fiscal Crisis of the State*. New York: St. Martin's Press.
- Pasinetti, L. (1997) "The Social Burden of High Interest Rates", in P. Arestis, G. Palma and M. Sawyer, (Editors), *Capital Controversy, Post-Keynesian Economics and the History of Economics: Essays in Honour of Geoff Harcourt*. Volume I, London: Routledge.
- Ricardo, D. (1817) *On the Principles of Political Economy and Taxation*. Edited by P. Sraffa with the collaboration of M. Dobb. London and Cambridge: Royal Economic Society and Cambridge University Press, 1995.
- Rochon, L-P. and M. Vernengo. (2003) "State Money and the Real World: Chartalism and its Discontents", *Journal of Post Keynesian Economics*, Volume 26, Number 1, pp. 57-67.
- Rotwein, E. (1955) "Introduction", in *David Hume, Writings on Economics*. Madison: University of Wisconsin Press, 1970.
- Schumpeter, J. (1918) "The Crisis of the Tax State", *International Economic Papers*, Volume 4, pp. 5-38, 1954.
- Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*.

- Edited by R. Campbell and A. Skinner.  
Indianapolis: Liberty Fund, 1981.
- Stiglitz, Joseph E. (2003) *The Roaring Nineties: A New History of the World's Most Prosperous Decade*. New York: W. W. Norton
- Tanzi, V. and L. Schuknecht. (2000) *Public Spending in the 20th Century*. Cambridge: Cambridge University Press.
- Tobin, J. (1965) "The Burden of the Public Debt: A Review Article", *Journal of Finance*, Volume 20, Number 4, pp. 679-682.
- Vernengo, M. (2004) "Political Economy and the Tax State: The Power of the Purse in Historical Perspective", in S. Durlabhji (Editor), *Power in Focus*. Bristol, UK: Wyndham Hall Press.
- Winch, D. (1998) "The Political Economy of Public Finance in the 'Long' Eighteenth Century", J. Maloney, (Editor), in *Debt and Deficits: An Historical Perspective*. Cheltenham, UK: Edward Elgar.

*Matias Vernengo*  
*Department of Economics*  
*University of Utah,*  
*Salt Lake City, USA*  
*matias.vernengo@economics.utah.edu*

## Human Capital and Knowledge

*Brian Chi-ang Lin*

### **Introduction**

Human capital and knowledge are frequently used interchangeably and broadly refer to people's acquired or innate abilities that are conducive to productivity and economic growth. Human capital is not only heterogeneous but, unlike physical capital, is also inseparable from an individual who possesses it. The concept of human capital is generally associated with the intangible attributes of people such as education and training, experience, skills, health, language (Lazear 1995, Pendakur & Pendakur 2002), or even attitudes toward work. Investment in human capital is seen as an improvement in people's productivity, which in turn can contribute to a higher level of economic growth for a country.

The present interpretation of the knowledge economy (or knowledge-based economy) focuses on the role of knowledge or human capital in long-run economic growth. Research on economic growth has been reawakened since the mid-1980s. According to the new growth theory developed by Romer (1986) and Lucas (1988), advancing knowledge is a crucial determinant of long-term growth. In recent years, discussions and investigations on the knowledge economy have intensified (see, e.g., Grossman & Helpman, 1991, Freeman & Polasky 1992, Jones 1995, OECD 1996, Atkinson & Court 1998, Aghion & Howitt 1998, Thurow 1999).

The significance of "human capital" and "knowledge", indeed, has been explored and elaborated by economists of different generations and groups, ranging from Adam Smith to contemporary Nobel laureates (such as Friedrich August von Hayek of the

Austrian school and Theodore W. Schultz, Gary S. Becker, and Robert E. Lucas of the Chicago school). In view of the long span and voluminous research on the subject, economists have classified types (or classes) of human capital and/or knowledge from different dimensions. Blaug (1976:829), for example, considered the following phenomena—health, education, job search, information retrieval, migration, and in-service training—as investments in human capital. According to Machlup (1962, 1980), knowledge can be classified into the following five types: (1) practical knowledge, (2) intellectual knowledge, (3) small-talk and pastime knowledge, (4) spiritual knowledge, and (5) unwanted knowledge. In addition, he classified knowledge production into six major knowledge industries and branches: (1) education, (2) research and development (R&D), (3) artistic creation and communication, (4) media of communication, (5) information services, and (6) information machines. According to some preliminary estimates provided by Machlup (1962), total knowledge-production for the U.S. in 1958 was \$136,436 million, with \$60,194 million in education, \$10,900 million in R&D, \$38,369 million in media of communication, \$8,922 million in information machines, and \$17,961 million in information services. The ratio of knowledge-production to adjusted GNP was almost 29 percent. It is anticipated that the share of knowledge-production in GDP has been increasing over the past several decades. According to the OECD (1996), more than half of GDP in the major OECD countries is now knowledge-based.

### **Early Thoughts and Intellectual Roots**

The concept of human capital is not novel in economic literature. An embryonic emphasis

on “human capital” or “knowledge” can be extended at least back to the early works of Adam Smith and David Ricardo (Kurz 1997, Nerdrum 1999:15-19). Smith envisaged the importance of education and skilled labor in his day and wrote in 1776 that “man educated at the expence of much labour and time ...may be compared to one of those expensive machines. .... The difference between the wages of skilled labour and those of common labour, is founded upon this principle” (Smith 1776:97-98).

Following Smith, Ricardo did not deny the existence of different kinds of labor (such as skilled and unskilled labor). There are, however, more references to Smith’s works than to Ricardo’s works. This probably has much to do with Ricardo’s vision of long-term economic development as a *stationary state*, which later gave economics the nickname *the dismal science*. Besides Smith and Ricardo, Thomas Robert Malthus argued that poverty could be resolved through institutional changes in enhancing human capital in the form of introducing universal education (Jensen 1999).

A sketch of the early contributors must, at least, include Irving Fisher’s capital theory and Joseph A. Schumpeter’s ideas of innovations. At first sight, Fisher did not seem to have much to do with the theory of human capital. Fisher, however, developed a heterogeneous capital concept that was functionally applied to literature on human capital thereafter (Nerdrum 1999:25-40). According to Fisher (1906:66), capital can be defined as “a quantity of wealth existing at an instant of time. ... Such a collection of wealth is, however, heterogeneous; ...They may, however, be reduced to a homogeneous mass by considering, not their kinds and quantities, but their values.” All types of capital (including human capital in a sense),

according to Fisher, that give rise to an income stream can be converted into a common standard, say, money. With further modification, it is not difficult to draw an analogy between physical and human capital. Fisher’s concept of capital has, as a result, thereafter been extended by the Chicago school to analyze human capital. According to Schumpeter (1942), under trustified capitalism, innovations were selected by large incumbent firms with market power that emphasized innovations as the key impetus to economic growth. Schumpeter (1942:82-83) argued that “in dealing with capitalism we are dealing with an evolutionary process...The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers’ goods, the new methods of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprise creates...This process of Creative Destruction is the essential fact about capitalism.” In Schumpeter’s view, the gap between firms creating technologies and firms appropriating them triggered periods of “creative destruction.”

### **Austrian Analysis of Knowledge**

The subject of human capital and knowledge has not only attracted individual economists but also different schools. The Austrian analysis of knowledge can be traced as far back as the early work of the founder of the Austrian School, Carl Menger, in his *Principles of Economics* published in 1871 (Baetjer 2000). The *knowledge problem*, however, was formally introduced and analyzed by Hayek in the 1930s and 1940s (Hayek 1937, 1945). Later, Hayek (1974) warned that economists pretended to know what was in practice not fully known or

measurable, and they inevitably risked giving false advice.

To Hayek, competition meant decentralized planning by heterogeneous individuals who possess differential knowledge. The best use of knowledge in society is to ensure that heterogeneous individuals with distinct plans can promptly apply their limited or partial knowledge to cooperate and/or compete with each other in the market. Hayek's notion of equilibrium, in this context, implies a specific situation in which all heterogeneous individuals' plans are synchronized. In addition, the interactions of all these heterogeneous individuals (best known as the market process or a *catallaxy*) can lead to the creation or discovery of new knowledge.

To paraphrase Hayek, Austrian economists recognize that the lack of (perfect) knowledge or human ignorance is constitutional. In essence, the economic problem is concerned with how heterogeneous individuals with limited knowledge carry out their actions and execute their plans over time through exchanges with each other. The market is an institution for the coordination, exchange, and utilization of the differential knowledge of individuals. People learn by doing through the market process and acquire new knowledge by interacting with others.

Machlup, a late president of the American Economic Association (AEA) and an eminent Austrian economist specializing in the subject of knowledge, followed the Austrian tradition and published some important works including *The Production and Distribution of Knowledge in the United States* (1962), *Knowledge and Knowledge Production* (1980), *The Branches of Learning* (1982) and *The Economics of Information and Human Capital* (1984). His unusual insights and works have become the

*de facto* standard, and have highlighted the importance of knowledge production for economic growth in modern economies which have stimulated subsequent research into the knowledge economy. From an Austrian perspective, the competitive market process, as illustrated above, has led to beneficial interaction among market participants. This process, over time, reduces ignorance for economic agents to manageable levels and promotes the discovery of knowledge that was not previously available. This dynamic process furthers the emergence of knowledge and could contribute to economic growth.

### **Human Capital and the Chicago School**

Although the Austrian school investigated the knowledge subject earlier and made a phenomenal contribution, it was the Chicago school economists, namely, T. W. Schultz, Gary Becker and Robert Lucas, who promoted the research domain of human capital to the frontier. T. W. Schultz, a native of South Dakota, delivered a presidential speech entitled "Investment in Human Capital" to the AEA in December 1960 and emphasized that the most unique feature of the US economic system was the growth in human capital. He described the important activities that improve human capabilities as follows:

"I shall concentrate on five major categories: (1) health facilities and services, broadly conceived to include all expenditures that affect the life expectancy, strength and stamina, and the vigor and vitality of a people; (2) on-the-job training, including old-style apprenticeship organized by firms; (3) formally organized education at the elementary, secondary, and higher levels; (4) study programs for adults that are not organized by firms, including extension

programs notably in agriculture; and (5) migration of individuals and families to adjust to changing job opportunities.” (Schultz 1961:9)

He concluded his speech by stating “The man without skills and knowledge [is] leaning terrifically against nothing” (Schultz 1961:16). Later, he published another short article entitled “Reflections on Investment in Man” in the *Journal of Political Economy* (JPE) in October 1962 and empirically examined the role of human capital as a source of economic growth. He identified that the growth in investment in man (i.e. educational capital) improved the quality of workers, which in turn became a major source of economic growth. In 1963, Schultz published another important book entitled *The Economic Value of Education*.

Immediately following Schultz’s 1962 article, a second article entitled “Investment in Human Capital: A Theoretical Analysis” written by Gary Becker and a third article entitled “On-the-Job Training: Costs, Returns, and Some Implications” written by Jacob Mincer appeared in the *Journal of Political Economy* in the October 1962 supplement on “Investment in Human Beings.” Both Becker (1962) and Mincer (1962) conducted in-depth analyses of on-the-job training. In 1964 Becker published *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, which became a classic book in the area of human capital. A special note should be reserved for Jacob Mincer, who maintained a close connection with both T. W. Schultz and Gary Becker at that time. Indeed, prior to both Schultz and Becker, Mincer employed the concept of human capital to explain the differentials of earnings and published a paper entitled “Investment in Human Capital and Personal Income Distribution” in the *Journal of*

*Political Economy* in 1958. In the 1980s, Lucas (1988) considered the *external effects* of human capital built on the concept of human capital developed by Schultz (1963) and Becker (1964). These effects are seen as spillovers from one person to another and to some extent contribute to the productivity of all factors of production. That is, human capital increases the productivity of both labor and physical capital. Lucas (1988:19) emphasized that “human capital accumulation is a *social* activity, involving *groups* of people in a way that has no counterpart in the accumulation of physical capital.”

Under the human-capital research program (the term used by Blaug 1976), some interesting issues with strong important policy implications emerge such as the phenomenon of human capital flight or the brain drain, which refers to the migration of the more highly-skilled labor to foreign countries to pursue higher rates of return. The brain drain issue, indeed, has drawn significant attention since the 1970s (see, for example, Bhagwati and Wilson 1989). Some recent studies have re-examined the brain drain issue and have considered human capital formation in the context of endogenous growth theory (Haque & Kim 1995, Beine *et al.* 2001, Stark 2004).

### **Empirical Analyses**

Since education is recognized as the most common type of human capital, it has by far drawn the most attention in empirical literature. Empirical studies such as Barro (1991) and Barro and Sala-i-Martin (1995), for example, have found a strong relationship between schooling and the growth rate of per capita GDP across countries. Bils and Klenow (2000) have recently pointed out that the impact of

schooling on growth explains less than one-third of the empirical cross-country relationship. The previous results documented by Barro (1991) and Barro and Sala-i-Martin (1995) can be best explained as partly reflecting the impact of schooling on growth. In this regard, they suggest that an important part of the relationship between schooling and growth may be due to omitted factors and such factors are subject to further study.

Foster and Rosenzweig (1996) employ longitudinal data of over 4,000 rural household in India covering the 1968-1981 period to investigate the complex relationships between schooling and economic growth. They find that technical change (in exploiting new high-yielding variety crop seeds) not only affects the returns to schooling but also leads to greater private investment in schooling. The central implication of their evidence indicates that the returns to investment in schooling and technical change are complementary to each other. The returns to investment in schooling will be higher if technical change is faster and the returns to investment in technical change will be higher if more schooling is available. Using data from the 1989 Taiwan Women and Family Survey, Parish and Willis (1993) investigate how parents choose to invest in sons' versus daughters' education and find that early-born female children in large families often marry early (rather than work) to solve the problem of tight family budgets. When older sisters do marry early, they can help increase the education of younger brothers and sisters. Zucker *et al.* (1998) investigate the close connection between intellectual human capital created by frontier research and the founding of firms in the high-tech industry. They conclude that the growth and diffusion of intellectual human capital was the

principal determinant of the growth and location of the U.S. biotechnology industry for the 1976-1989 period.

Haveman *et al.* (2003) have recently developed an indicator, earnings capacity (EC), to measure the U.S. human capital stock value from 1975 to 2000 and the extent to which that capital has been utilized. The EC measure is a potential measure and reveals what annual earnings in the US would be if all working-age population employed their human capital at its capacity. Then, the potential EC can be compared to the actual utilization of human capital. According to their estimates, the annual rental value of the human capital stock in the U.S. was about \$6.26 trillion in 2000, and it had grown by almost 74 percent since 1975. In 2000, the average working-age individual had over \$39,000 of EC, with the average male EC at about \$46,500 and the average female EC at about \$32,000. The EC gap between males and females has substantially narrowed over the 1975-2000 period. Furthermore, their estimates indicate that utilization of human capital has fluctuated from a low of 60 percent in 1982 to a high of 72 percent in 2000.

### **Human Capital, Poverty, and Institutions**

Although utilization of human capital has generally increased (especially attributable to women's human capital) in the U.S. over the 1975-2000 period, Haveman *et al* (2003) did not further investigate the relationship between human capital and poverty. It is well known that the burden of poverty has heavily fallen on women and children, and this phenomenon is particularly significant for developing countries. Furthermore, people who suffer from poverty frequently possess some characteristics such as bad health conditions. An improvement in people's health, from the perspective of the

mainstream exposition of human capital, is vital for resolving the poverty problem. Poverty, however, has become a chronic phenomenon and has continued to prevail in our human societies. In fact, the poverty issue can only be better understood from a broad institutional dimension. Borjas (1995), for example, examines human capital from an ethnic perspective and shows the empirical evidence between ethnic neighborhoods and the spillover effect of “ethnic capital.” His study clearly indicates that ethnic skill differentials persist from generation to generation and the ethnic-capital effect comes into play through the channel that low-income groups cluster in low-income ethnic neighborhoods. Intergenerational mobility is greatly affected by neighborhood effects.

A stylized fact on earth is that many poor people relentlessly die of hunger and disease every day. Locally and globally, our human societies have been continuously characterized by the scenario of pursuing a growing economy but being accompanied by rising wealth inequalities. In this regard, Kenneth E. Boulding, a late AEA president and a well-known institutionalist, emphasized that the knowledge of economics does not include the economics of knowledge and stated that “one area where economists have a good deal to be humble about is in the field of economic development of poor countries” (Boulding 1966, p. 11).

Gunnar Myrdal, a 1974 Nobel laureate and also a well-known institutionalist, investigated the South Asia region (including Pakistan, India, Indonesia, Burma, Philippines, Thailand, Ceylon, Malaya, and sometimes South Vietnam, Cambodia, and Laos as well) and published a three-volume book entitled *Asian Drama: An Inquiry Into the Poverty of Nations* in

1968. Almost five decades later, some of his penetrating insights still deserve constant attention. The entire volume III of *Asian Drama, Problems of Population Quality*, was devoted to the study of health, education, and the social system as a whole, with an application of his analysis to government action.

From a holistic perspective, the socio-economic conditions in each South Asian country can be classified into six broad categories, i.e. (1) output and incomes; (2) conditions of production; (3) levels of living; (4) attitudes toward life and work; (5) institutions; and (6) policies (Myrdal 1968, pp. 1859-1864). The first three categories represent “economic factors” and categories 4 and 5 represent “non-economic factors.” Category 6 is a mixture and can be considered to belong to the “economic factors” when the policies purpose is to induce changes in the first three conditions. In the social system all the aforementioned conditions are causally interrelated and economic conditions do not have precedence over the others. From the viewpoint of development, a change of condition in the direction of greater desirability can be considered as a change “upwards.”

Myrdal pointed out that the South Asian people have not only been insufficiently educated but also been miseducated to a great extent (due to their past colonial rule). Thus, educational reforms needed in these now independent countries are far more than the popular suggestion of increased “investment in education.” The main reason is that “existing educational establishments are part of a larger institutional system, which includes social stratification; and this system is supported by people’s attitudes, which themselves have been molded by the institutions” (Myrdal 1968:1649). Influential vested interests in these countries have been



embedded in the educational and institutional systems and they resist or warp reform policies.

According to Boulding and Myrdal, it is clear that economists need to develop substantial knowledge towards the poor to resolve the poverty problem. Poverty is frequently entangled with the phenomena such as contagious diseases and criminal activities, which cause great negative impact on other individuals and communities. Reducing poverty, in a sense, possesses the property of public good if poverty reduction induces greater public health and security. From a global village perspective, the reduction of poverty has become even more significant if it contributes to disease eradication and global peace, both of which are considered as important global (or international) public goods.

### **International Public Goods (IPGs) and Knowledge Spillovers**

Knowledge, as previously illustrated for the role of reducing poverty, is also conceptually a type of IPG. The significance of IPGs has recently given rise to intense analyses in the literature (see, for example, Kaul *et al.* 1999 and Ferroni and Mody 2002). A pure IPG, in principle, can generate benefits that spill over borders, regions, ethnic groups, and generations. Morrissey *et al.* (2002) classify IPGs into five categories: environment, health, knowledge, security, and governance.

Types of IPGs as diverse as internet securities, financial market stability, biodiversity preservation, and knowledge (of best development practices) can be framed in terms of the so-called *weakest link*, *weaker link*, *better shot*, and *best shot* public goods (Sandler 1998 and Arce M. 2004). The introduction of these adjectives characterize the variety of ways in which the

international collective action is translated into the provision of public goods. For instance, the smallest effort or contribution uniquely determines the public good level for a *weakest link* technology, while the largest effort or contribution uniquely determines the public good level for a *best shot* technology. Sandler (1998, p. 232) pointed out that some of the most worrisome public good challenges facing human beings adhere to the best-shot technology (such as finding a cure for avian flu). It is quite conceivable that the provision of global public goods has remained under-provided (in the absence of an international government) and faced with a global governance challenge. In his presidential address entitled “International Public Goods without International Government” to the AEA, Kindleberger (1986:11) commented that “the system should be run at all times by rules, including regimes, not people.”

Among the aforementioned public goods, the benefit that knowledge (created by the innovative firms) spills over to other firms is referred to as “knowledge spillovers,” which are considered as “engine of endogenous economic growth.” Many empirical studies have examined knowledge spillovers from a global dimension and provided different results. Using pooled time series cross section data for 21 OECD countries and Israel, Coe and Helpman (1995) discover that foreign R&D capital stocks have a significant impact on domestic productivity, indicating the existence of international knowledge spillovers. Using firm-level data for the U.S. and Japan, Branstetter (2001) only find strong evidence of *intranational* knowledge spillovers. These empirical findings are quite important for understanding the channels and barriers of knowledge spillovers across nations. Furthermore, their implications are

significant for furthering policy formation of knowledge as an IPG.

Indeed, the spread of knowledge or knowledge spillovers in a global society is greatly involved with the increase in traded amounts of knowledge-related goods and services. As property rights can be attached to identifiable pieces of knowledge, the intellectual property rights (IPRs) system (such as the patent right, copyright, semiconductor chip protection, and trademark protection for the U.S.) has become prevalent and dominant in controlling access to knowledge, the spread of knowledge, and also the trading of knowledge-related goods and services.

### **Knowledge and Trade-related Intellectual Property Rights (TRIPS)**

In recent years, the idea of IPRs and the enforcement of trade-related aspects of intellectual property rights (TRIPS), which integrated IPRs into global trade, have attracted increased attention and debate. Under the insistence of the U.S. Trade Representative, the GATT agreement under the 1993 Uruguay Round concluded provisions in relation to TRIPS. TRIPS is an international agreement that operates under the umbrella of the WTO, which was established in 1995 and succeeded the GATT as the forum for international trade negotiations. TRIPS establishes minimum levels of protection in relation to IPRs such as copyright and drug patents that each government has to give to the intellectual property of fellow WTO members.

Proponents of the IPRs, on the one hand, believe that the current system can not only protect the process of invention and innovation but can also provide economic incentives for creative activities. Opponents, on the other hand, generally hold the view that the protection of IPRs has been

implemented at the expense of the public knowledge domain. Richards (2002) examined justificatory arguments in defense of IPRs based on the philosophical writings of Locke, Hegel and Bentham and found that their defenses for private property rights do not hold up well when applied to intellectual property. Boldrin and Levine (2002) thought that intellectual property might be better referred to as “intellectual monopoly” and argued that the extent of current copyrights has been excessive, with a similar conclusion also having been reached by Lessig (2001). Hui and Png (2002) studied the impact of a change in the U.S. copyright law in 1998 in relation to movie production and found that the Sonny Bono Act has only had a small impact on new creative activity. To facilitate the production and use of knowledge, Joseph Stiglitz (1999), 2001 Nobel laureate, argued that an international intellectual property system must balance a variety of concerns such as dynamic efficiency (derived from innovative activity), static efficiency (associated with the utilization of knowledge) and the use of the global knowledge commons. As each innovative activity currently makes free use of the global commons of (pre-existing) knowledge, the international community could claim the right to charge a fee for using it.

TRIPS has so far become the most important backdrop for global governance in the advance and spread of knowledge. Ideally, TRIPS not only aims to promote creativity and innovation, but also allows a built-in flexibility among countries to adapt them to their own public purposes (see Article 7 & Article 8.1). In practice, there has been resistance in the developing world where it has been argued that their interests were under-represented (and IPRs primarily work for the interests of developed

countries, and particularly for the U.S.). According to Ryan (1998), TRIPS was established on behalf of the U.S. multinational companies such as Pfizer Chemical and IBM. Shaffer (2004) notes structural weakness in the ability of developing countries to participate in the WTO judicial process and offers suggestions for protecting their interests. Lamy (2004) views TRIPS as a flexible instrument and only one part of a global governance mechanism and points out that IPRs can become a source of growth for developing countries only if sound economic management and the right type of flanking policies are considered simultaneously. Apparently, further studies on property rights laws and institutions are required to evaluate the IPR system.

### **Indigenous Knowledge and Sustainable Development**

The efforts to comprehend the importance of *human capital* or *knowledge* are by no means limited to economists. To further expand the global knowledge commons, a serious acknowledgement and thorough understanding of a unique intellectual knowledge system – indigenous knowledge – seems to be indispensable. The recognition of and emphasis on indigenous knowledge has been intensified by recent developments in environmentalism and feminism (Jacobs 1994). Indigenous knowledge usually refers to unwritten knowledge preserved locally in oral traditions and has been increasingly recognized as critical for sustainable development (see, for example, Brokensha *et al.*, 1980 and Fernando 2003). A central argument is that sustainable development must be embedded in indigenous knowledge systems, and ignorance of the systems will certainly lead to failure in development. Weatherford (1994) argued that the

indigenous people may be the only people capable of salvaging modern civilization. As indigenous people have gradually lost their cultural identities, we are actually losing indigenous knowledge, our connections to the past, and jeopardizing our future.

According to Fernando (2003:58), it would be far more productive to regard indigenous knowledge as “a social phenomenon produced within a specific social, economic, and political context and, thereafter, proceed to analyze the relevance of such meanings and the institutional and power relations embedded in them for achieving the goals of sustainable development.” Indigenous knowledge is seen to be particularly abundant in the knowledge of the environment or nature and, therefore, critically important to sustainable resource use and balanced development (Brokensha *et al* 1980). Take two of Taiwan’s indigenous tribes, Thao and Tsou, for example. The Thao are Taiwan’s smallest ethnic group with a population of about 300 people and live in the surrounding area of Sun Moon Lake. The Tsou population is approximately 7,000 people and most of them live in the Alishan area. Alishan and Sun Moon Lake are two of most famous scenic areas for enjoying natural beauty in Taiwan. Without the Thao and Tsou’s indigenous wisdom, the natural environment of Sun Moon Lake and Alishan cannot be well preserved until today.

The following is another real example of indigenous knowledge. The Onge people make up an aboriginal tribe with a population of less than 100 people. They live in the Indian Nicobar and Andaman islands in the Bay of Bengal. They possess indigenous knowledge encompassing medicine, biology, and nature (Norchi 2000). They rely on their knowledge of a specific plant to treat fever and

gastrointestinal disorders. This same plant is also effective in dealing with malaria and, consequently, the Onge people no longer suffer from malaria. Now, scientists have been conducting tests on the medicinal plants used by the Onge people and pharmaceutical companies have also shown an interest in reaching an agreement with the patentee. (Note: New cures for malaria could be regarded as a *best shot* international public good.) In addition, all 96 Onge people survived the tsunami in December 2004. The tsunami, however, has left around 300,000 people missing or dead in or around the Indian Ocean, including residents of the Onge people's nearby town of Hut Bay. Before the tsunami, the Onge people fled into the jungle for protection after they found that the water in the creek had suddenly run out to sea.

The work of protecting and promoting indigenous knowledge, however, has been challenging. The following urgent issues need to be examined and evaluated. First, it is important to conduct research on how people can preserve the natural environment of the indigenous people as indigenous knowledge and their natural habitats go hand in hand. Second, the existing intellectual property system still remains controversial. Thus, whether it can be applied or extended to cover indigenous knowledge needs in-depth research. Third, a unique institutional system for the advancement of indigenous knowledge needs to be developed. The US Tribal Colleges, indeed, were created in the late 1960s and the early 1970s in response to the higher education needs of American Indians. There are currently over 30 Tribal Colleges located in Indian reservations around the US. Tribal Colleges are different from standard community colleges or mainstream 4-year colleges in their cultural identities, and are relatively involved in a

broad range of community services – including education, counseling, and economic development initiatives – that are specifically focused on communities that would otherwise be isolated from such resources. So far, the Tribal Colleges have not drawn enough attention from either academia or the general public and have been persistently under financial pressure.

### **Sustainable Perspective on the Knowledge Economy**

Finally, it is time to seriously reconsider the knowledge issue from the perspective of our position in human history. Clearly, knowledge is not only critical for economic growth but also for poverty alleviation and environmental conservation. In short, knowledge is critical for a sustainable society. In this respect, John Stuart Mill's concept of *the stationary state* is in line with contemporary analysis of sustainable development and is worth further review.

O'Connor (1997) investigates Mill's concepts of a private property-based liberal society as well as a stationary-state society and argues that the writings of Mill represent a prototype for ideals of a "sustainable development." Winch (2004:111) points out that Mill is one of the earliest green thinkers and his "defense of a zero-growth society conveys the substance of his environmentalist concerns." Mill's virtuous stationary-state (zero-growth) society, according to Winch (2004:122), is "a continuous state of dynamic equilibrium" in which all improvements in new technologies can be redirected towards redistribution of wealth and the promotion of life quality.

Lin (2003, 2005) has further argued that the rich communities (nations) have tended to waste resources, whereas the poor communities (nations) have tended to destroy resources. Due to severe wealth

inequalities all over the world and limited resources on earth, the global community has become less and less sustainable. To maintain a sustainable society with an efficient use of resources, it is necessary to achieve a more equitable distribution of wealth. In this regard, Lin (2003, 2005) has restated that Mill's concept of *the stationary state* is conceptually consistent with modern exposition of sustainable development.

Although greatly influenced by David Ricardo, Mill's *stationary state* was not the dismal one David Ricardo visualized. Mill took a different view of his desirable society and outlined his desires for a good future. Indeed, in his chapter on *the stationary state*, in which he discussed the long-run tendencies of the economy, he said:

“But the best state for human nature is that in which, while no one is poor, no one desires to be richer, nor has any reason to fear being thrust back by the efforts of others to push themselves forward. ...There would be as much scope as ever for all kinds of mental culture, and moral and social progress; as much room for improving the Art of Living, and much more likelihood of its being improved, when minds ceased to be engrossed by the art of getting on.” (Mill 1848:748-51)

Looking at the economic and social conditions of his time, Mill felt that the mass of society was bypassed by the materialistic development of the Industrial Revolution and wondered whether a country with a growing economy was a desirable living place. He envisioned that *the stationary state* would result in an improvement in the art of living and emphasized that “only in the backward countries of the world that increased production is still an important object: in those most advanced, what is

economically needed is a better distribution” (Mill 1848, p. 749).

Thus, Mill's *stationary state* might be narrowly interpreted as a society with *no (or limited) growth in physical output*. Alternatively, it should be best understood as a society with unlimited growth in mental culture and improvements in economic equality (by means of wealth redistribution). Mill's perspective, in fact, is very much closed to the ethical-utopian perspective of sustainable development. The ethical-utopian perspective of sustainable development emphasizes the emergence of new individual value criteria and new social objectives. Furthermore, it encourages altruistic behavior on the part of modern citizens in contrast to individual egoistic behavior (Bergh 1996:59).

From the perspective of the mainstream literature, an improvement in human capital is seen as crucial for enhancing the *competitiveness* of a single person or country. To date, we have begun to inquire into the possibility of a global *sustainability* from an overlapping-generations perspective. Fundamental to this holistic perspective is the recognition that human generations are interrelated and ought to be examined as an integrated whole. This ongoing progress, indeed, implies our intellectual breakthrough to some extent and also reveals our common desires for a sustainable future. Undoubtedly, our human generations will continue to develop new values, beliefs, and knowledge compatible with the goal of a sustainable future. What I have offered here, therefore, are only some very basic prolegomena on the subject.

**Acknowledgments:** This article was originally completed during the author's visit on a Fulbright scholarship to Augustana College, South Dakota, USA. The author

gratefully acknowledges the auspices of Augustana College and is indebted to Reynold Nesiba, Brian Eggleston, David Sorenson, and Donn Grinager for their generous support. The author also wishes to thank Phil O'Hara, Yih-Chyi Chuang, and anonymous referees for helpful comments and the Fulbright program, administered by the Institute of International Education and sponsored by the United States Department of State, which has helped with the creation and spread of knowledge, an important type of international public good.

### Internet Sites

Commission on Intellectual Property.

[www.iprcommission.org](http://www.iprcommission.org)

Global Public Goods.

[www.earthsummit2002.org/es/issues/GPG/gpg.htm](http://www.earthsummit2002.org/es/issues/GPG/gpg.htm)

Indigenous Knowledge Pages. [www.ik-pages.net](http://www.ik-pages.net)

Trade-Related Intellectual Property Rights.

[www.wto.org/english/tratop\\_e/trips\\_e/trips\\_e.htm](http://www.wto.org/english/tratop_e/trips_e/trips_e.htm)

### Selected References

Aghion, Philippe and Peter Howitt. (1998) *Endogenous Growth Theory*. MIT Press, Cambridge, MA.

Arce M., Daniel G. (2004) "Asymmetric Leadership and International Public Goods" *Public Finance Review*, Volume 32, Number 5, September, pp. 528-558.

Atkinson, Robert D. and Randolph H. Court. (1998) *The New Economy Index: Understanding America's Economic Transformation*. Progressive Policy Institute, Washington, DC.

Baetjer, Howard, Jr. (2000) "Capital as Embodied Knowledge: Some Implications for the Theory of Economic Growth." *Review of Austrian Economics*,

Volume 13, Number 2, September, pp. 147-174.

Barro, Robert J. (1991) "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics*, Volume 106, Number 2, May, pp. 407-443.

Barro, Robert J. and Xavier Sala-i-Martin. (1995) *Economic Growth*. McGraw-Hill, New York.

Becker, Gary. (1962) "Investment in Human Capital: A Theoretical Analysis." *Journal of Political Economy* (Supplement), Volume 70, Number 5, October, pp. 9-49.

Becker, Gary. (1964) *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. National Bureau of Economic Research, New York.

Beine, Michel; Frédéric Docquier and Hillel Rapoport. (2001) "Brain Drain and Economic Growth: Theory and Evidence." *Journal of Development Economics*, Volume 64, Number 1, February, pp. 275-289.

Bergh, Jeroen C.J.M. van den. (1996) *Ecological Economics and Sustainable Development: Theory, Methods and Applications*. Edward Elgar, Cheltenham, UK.

Bhagwati, J. and J.D. Wilson. (1989) *Income Taxation and International Mobility*. MIT Press, Cambridge, MA.

Bils, Mark and Peter Klenow. (2000) "Does Schooling Causes Growth?" *American Economic Review*, Volume 90, Number 5, December, pp. 1160-1183.

Blaug, Mark. (1976) "The Empirical Status of Human Capital Theory: A Slightly Jaundiced Survey." *Journal of Economic Literature*, Volume 14, Number 3, September, pp. 827-855.

Boldrin, Michele and David K. Levine. (2002) "The Case Against Intellectual

- Property.” *American Economic Review* (Papers and Proceedings), Volume 92, Number 2, May, pp. 209-212.
- Borjas, George J. (1995) “Ethnicity, Neighborhoods, and Human-Capital Externalities.” *American Economic Review*, Volume 85, Number 3, June, pp. 365-390.
- Boulding, Kenneth E. (1966) “The Economics of Knowledge and the Knowledge of Economics.” *American Economic Review*, Volume 56, Number 2, May, pp. 1-13.
- Branstetter, Lee G. (2001) “Are Knowledge Spillovers International or Intranational in Scope? Microeconomic Evidence from the U.S. and Japan.” *Journal of International Economics*, Volume 53, Number 1, February, pp. 53-79.
- Brokensha, D.; D. Warren and O. Werner (1980) (Editors) *Indigenous Knowledge Systems and Development*. University Press of America, Lanham, MD.
- Coe, David T. and Elhanan Helpman. (1995) “International R&D Spillovers.” *European Economic Review*, Volume 39, Number 5, May, pp. 859-887.
- Fernando, Jude L. (2003) “NGOs and Production of Indigenous Knowledge under the Condition of Postmodernity.” *Annals of the American Academy of Political and Social Sciences*, Volume 590, November, pp. 54-72.
- Ferroni, Marco A. and Ashoka Mody. (2002) (Editors) *International Public Goods: Incentives, Measurement, and Financing*. Kluwer Academic Publishers, Boston.
- Fisher, Irving. (1906) *The Nature of Capital and Income*. Augustus M. Kelley, New York, 1965.
- Foster, Andrew D. and Mark R. Rosenzweig. (1996) “Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution.” *American Economic Review*, Volume 86, Number 4, September, pp. 931-953.
- Freeman, Scott and Stephen Polasky. (1992) “Knowledge-Based Growth.” *Journal of Monetary Economics*, Volume 30, Number 1, October, pp. 3-24.
- Grossman, Gene M. and Elhanan Helpman. (1991) *Innovation and Growth in the Global Economy*. The MIT Press, Cambridge, MA.
- Haque, Nadeem U. and Se-Jik Kim. (1995) “Human Capital Flight: Impact of Migration on Income and Growth.” *IMF Staff Papers*, Volume 42, Number 3, September, pp. 577-607.
- Haveman, Robert H.; Andrew Bershadker and Jonathan A. Schwabish. (2003) *Human Capital in the United States from 1975 to 2000: Patterns of Growth and Utilization*. W. E. Upjohn Institute for Employment Research, Kalamazoo, MI.
- Hayek, Friedrich August von. (1937) “Economics and Knowledge.” *Economica*, Volume 4, February, pp. 33-54.
- Hayek, Friedrich August von. (1945) “The Use of Knowledge in Society.” *American Economic Review*, Volume 35, Number 4, September, pp. 519-530.
- Hayek, Friedrich August von. (1974) “The Pretence of Knowledge.” The Nobel Prize Lecture, December 11.
- Hui, Kai-Lung and I.P.L. Png. (2002) “On the Supply of Creative Work: Evidence from the Movies.” *American Economic Review* (Papers and Proceedings), Volume 92, Number 2, May, pp. 217-220.
- Jacobs, Jane M. (1994) “Earth Honoring: Western Desires and Indigenous Knowledges,” in Blunt, Alison and Gillian Rose (Editors), *Writing Women*

- and Space: Colonial and Postcolonial Geographies*. Guilford Press, New York, pp. 169-196.
- Jensen, Hans. E. (1999) "The Development of T.R. Malthus's Institutional Approach to the Cure of Poverty: from Punishment of the Poor to Investment in Their Human Capital." *Review of Social Economy*, Volume 57, Number 4, December, pp. 450-465.
- Jones, Charles I. (1995) "R&D-Based Models of Economic Growth." *Journal of Political Economy*, Volume 103, Number 4, August, pp. 759-784.
- Kaul, Inge; Isabelle Grunberg and Marc A. Stern (eds). (1999) *Global Public Goods: International Cooperation in the 21st Century*. Oxford University Press, New York.
- Kindleberger, Charles P. (1986) "International Public Goods without International Government." *American Economic Review*, Volume 76, Number 1, March, pp. 1-13.
- Kurz, Heinz D. (1997) "What Could the 'New' Growth Theory Teach Smith or Ricardo?" *Economic Issues*, Volume 2, Part 2, September, pp. 1-20.
- Lamy, Pascal. (2004) "Trade-Related Aspects of Intellectual Property Rights – Ten Years Later." *Journal of World Trade*, Volume 38, Number 6, December, pp. 923-934.
- Lazear, Edward P. (1995) "Culture and Language." NBER Working Paper 5249. National Bureau of Economic Research, Cambridge, MA.
- Lessig, Lawrence. (2001) *The Future of Ideas: the Fate of the Commons in a Connected World*. Random House, New York.
- Lin, Brian C.-A. (2003) "An Exploration of John Stuart Mill's Concepts of Sustainable Development and Fiscal Policy." National Science Council Project Report (NSC 91-2415-H-004-011), Taipei, Taiwan.
- Lin, Brian C.-A. (2005) "A Sustainable Perspective on the Knowledge Economy: A Critique of Austrian and Mainstream Views." Mimeo, National Chengchi University, Taipei, Taiwan.
- Lucas, Robert E., Jr. (1988) "On the Mechanics of Economic Development." *Journal of Monetary Economics*, Volume 22, Number 1, July, pp. 3-42.
- Machlup, Fritz. (1962) *The Production and Distribution of Knowledge in the United States*. Princeton University Press, Princeton, NJ.
- Machlup, Fritz. (1980) *Knowledge and Knowledge Production* (Volume I of *Knowledge: Its Creation, Distribution, and Economic Significance*). Princeton University Press, Princeton, NJ.
- Machlup, Fritz. (1982) *The Branches of Learning* (Volume II of *Knowledge: Its Creation, Distribution, and Economic Significance*). Princeton University Press, Princeton, NJ.
- Machlup, Fritz. (1984) *The Economics of Information and Human Capital* (Volume III of *Knowledge: Its Creation, Distribution, and Economic Significance*). Princeton University Press, Princeton, NJ.
- Mill, John Stuart. (1848) *Principles of Political Economy with Some of Their Applications to Social Philosophy* (Edited with Introduction by W. J. Ashley). Augustus M. Kelley, New York, 1965.
- Mincer, Jacob. (1958) "Investment in Human Capital and Personal Income Distribution." *Journal of Political Economy*, Volume 66, Number 4, August, pp. 281-302.
- Mincer, Jacob. (1962) "On-the-Job Training: Costs, Returns, and Some Implications."



- Journal of Political Economy* (Supplement), Volume 70, Number 5, October, pp. 50-79.
- Morrissey, Oliver; Dirk Willem te Velde and Adrian Hewitt. (2002) "Defining International Public Goods: Conceptual Issues," in Ferroni, Marco A. and Ashoka Mody (eds), *International Public Goods: Incentives, Measurement, and Financing*. Kluwer Academic Publishers, Boston, pp. 31-46.
- Myrdal, Gunnar. (1968) *Asian Drama: An Inquiry Into the Poverty of Nations* (3 Volumes). Pantheon, New York.
- Nerdrum, Lars. (1999) *The Economics of Human Capital: A Theoretical Analysis Illustrated Empirically*. Scandinavian University Press, Oslo.
- Norchi, Charles H. (2000) "Indigenous Knowledge as Intellectual Property." *Policy Sciences*, Volume 33, Number 3-4, December, pp. 387-398.
- O'Connor, Martin. (1997) "John Stuart Mill's Utilitarianism and the Social Ethics of Sustainable Development." *European Journal of the History of Economic Thought*, Volume 4, Number 3, Autumn, pp. 478-506.
- OECD. (1996) *The Knowledge-Based Economy*, Paris.
- Parish, William L. and Robert J. Willis. (1993) "Daughters, Education, and Family Budgets: Taiwan Experiences." *Journal of Human Resources*, Volume 28, Number 4, Fall, pp. 863-898.
- Pendakur, Krishina and Ravi Pendakur. (2002) "Language as Both Human Capital and Ethnicity." *International Migration Review*, Volume 36, Number 1, Spring, pp. 147-177.
- Richards, Donald G. (2002) "The Ideology of Intellectual Property Rights in the International Economy." *Review of Social Economy*, Volume 60, Number 4, December, pp. 521-541.
- Romer, Paul M. (1986) "Increasing Returns and Long-Run Growth." *Journal of Political Economy*, Volume 94, Number 5, October, pp. 1002-1037.
- Ryan, Michael P. (1998) *Knowledge Diplomacy: Global Competition and the Politics of Intellectual Property*. Brookings Institution Press, Washington, D.C.
- Sandler, Todd. (1998) "Global and Regional Public Goods: A Prognosis for Collective Action." *Fiscal Studies*, Volume 19, Number 3, August, pp. 221-247.
- Schultz, Theodore W. (1961) "Investment in Human Capital." *American Economic Review*, Volume 51, Number 1, March, pp. 1-17.
- Schultz, Theodore W. (1962) "Reflections on Investment in Man." *Journal of Political Economy* (Supplement), Volume 70, Number 5, October, pp. 1-8.
- Schultz, Theodore W. (1963) *The Economic Value of Education*. Columbia University Press, New York.
- Schumpeter, Joseph A. (1941) *Capitalism, Socialism, and Democracy* (first published in 1942, with a new introduction by Tom Bottomore), 1st Harper Colophon edition. Harper & Row, New York, 1975.
- Shaffer, Gregory. (2004) "Recognizing Public Goods in WTO Dispute Settlement: Who Participates? Who Decides? The Case of TRIPS and Pharmaceutical Patent Protection." *Journal of International Economic Law*, Volume 7, Number 2, June, pp. 459-482.
- Smith, Adam. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, A Selected Edition (Edited with Introduction and Commentary by

- Kathryn Sutherland). Oxford University Press, New York, 1993.
- Stark, Oded. (2004) "Rethinking the Brain Drain." *World Development*, Volume 32, Number 1, January, pp. 15-22.
- Stiglitz, Joseph E. (1999) "Knowledge as a Global Public Good," in Kaul, Inge; Isabelle Grunberg and Marc A. Stern (eds), *Global Public Goods: International Cooperation in the 21st Century*. Oxford University Press, New York, pp. 308-325.
- Thurow, Lester C. (1999) *Building Wealth: The New Rules for Individuals, Companies, and Nations in a Knowledge-Based Economy*. HarperCollins, New York.
- Weatherford, Jack M. (1994) *Savages and Civilization: Who Will Survive?* Crown Publishers, New York.
- Winch, Donald. (2004) "Thinking Green, Nineteenth-Century Style: John Stuart Mill and John Ruskin," in Mark Bevir and Frank Trentmann (Editors), *Markets in Historical Contexts: Ideas and Politics in the Modern World*. Cambridge University Press, New York, pp. 105-128.
- Zucker, Lynne G.; Michael R. Darby and Marilyn B. Brewer. (1998) "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review*, Volume 88, Number 1, March, pp. 290-306.

*Brian Chi-ang Lin*  
*Department of Public Finance*  
*National Chengchi University*  
*MuJa, Taipei, Taiwan*  
*calin@nccu.edu.tw*

## Income Inequality and the Distribution of Power

*Oren M. Levin-Waldman*

### **Introduction**

Income inequality can very simply be defined in terms of a gap between the haves and the have-nots. When applied to a country's income distribution it means that those at the top of the distribution hold a greater share than those in the middle and those at the bottom. Over the last couple of decades there has been a considerable increase in income inequality in much of the western industrialized world. Generally speaking, income inequality in the United States has been greater than other industrialized countries as measured by Gini coefficients, the ratio of the top quintile of family income to the bottom quintile, and the share of wealth held by those at the top relative to that held by those at the bottom. And yet, inequality in underdeveloped and developing countries lacking in sophisticated welfare state programs and where economic reforms, including land reforms, have not occurred, tends to be much higher than in the industrialized ones, including the United States.

Inequality, however, may be greater in countries that are in the process of transitioning to market economies. China serves as a good example where inequality was relatively low during the first years of Communist rule, especially when land reform was introduced. But it rose precipitously during what has been referred to as the Great Leap Forward and Great Famine from 1957-1966, reaching an all-time high in 1966. Though it declined afterwards, it began to rise again in 1987 after the government began undertaking market reforms (Kambur & Zhang 2001; Benjamin

et al 2005). Wang Xiaolu (2006) notes that the Gini coefficient was 32.0 in 1980, dropped to 25.7 during the initial stage of economic reform between 1980-1984, but increased to 35.5 in 1990 and then 44.7 in 2001.

Although free markets may be viewed as the source of inequality, the question remains: why is it greater in some countries than others? The simple answer is that the politics of a nation and the public policies that it pursues has much to do with the level of inequality. Short of adopting socialist policies, there is no way to really end it. But inequality can be lessened when a country seeks to pursue policies that bolster the least advantaged members of society.

This essay will explore trends in inequality and the various measures of it, and will also look at its impact on the distribution of power. While income inequality is a problem inherent to all capitalist economies, it tends to be greater in those countries with less of a social safety net. Indeed, it has risen in those countries where as a matter of deliberate public policy the social wage has decreased.

Income inequality, however, is more than a matter of distribution; rather it has serious repercussions for democratic theory because to the extent that it results in some having more resources than others it seriously affects the distribution of political power and ultimately the outcomes of a democratic political process. The privileged participate more than others and tend to be better organized to press their demands on government (Jacobs & Skocpol 2005). The reality, however, is that one cannot have true democracy amidst rising income inequality because in the end not all in the political process are equal.

## Measuring Income Inequality

Measuring income inequality is really no easy task. Four widely used measures of income inequality are the Gini coefficient, the Atkinson inequality index, the Theil inequality index, and the coefficient of variation.

The gini coefficient is a summary statistic ranging from 0 when all individuals are equal to 1 where there is complete inequality, and it tends to be very sensitive to changes around the median. Developed by the Italian statistician Corrado Gini in 1912, the index represents a percentage which is equal to the Gini coefficient multiplied by 100. A coefficient of .70, for example, would suggest extreme inequality, whereas a coefficient of .13 would suggest relatively low levels of inequality. How do countries compare using this index? Whereas most developed European nations have Gini coefficients between .24 and .36, the United States tends to have a coefficient above .4. And many Central and South American countries, as well as African countries, have Gini coefficients close to .60 and in some cases exceed it. The Gini coefficient, however, may not be the best measure when comparing large countries to small countries. Moreover, because benefit systems vary from country to country, comparisons of inequality between countries may also be difficult using this measure.

The Atkinson index explicitly incorporates normative judgments about social welfare and is derived from calculating an equity-sensitive average income based on a per capita income that would result in the total welfare being exactly equal to the total welfare generated by the actual income distribution if everybody enjoyed the same per capita income. The Atkinson index is sensitive to inequality changes in the lowest part of the

income distribution. Therefore, as equity sensitive income rises, more social weight is attached to income transfers at the lower end of the distribution. The Atkinson index in economics is specifically used to quantify income inequality. And more specifically, it is used to gauge movements in different segments of the income distribution.

The Theil inequality index, derived from the econometrician Henri Theil, ranges from 0 to infinity with higher values representing greater equality. It is essentially a mathematical formula:

$$(1) \quad T = \frac{1}{N} \sum_{i=1}^N X_i \left( \frac{x_i}{x} \ln \frac{x_i}{x} \right)$$

In the formula,  $x_i$  is the income of the  $i$ th person,  $x$  is the mean income, and  $N$  is the number of people. The first term inside the sum represents the individual's share of aggregate income, with the second term representing that person's income relative to the mean. When everybody has the same mean, the index is equal to zero, but when one person has all the income, the index is equal to  $\ln N$ . An advantage to the Theil index over say the Gini coefficient is that it is the weighted sum of inequality within subgroups. As an example, inequality in the U.S. would be the sum of each state's inequality weighted by the state's income relative to the entire country.

And the coefficient of variation is simply a statistical measure of the deviation from the mean. It measures dispersion of a probability distribution, defined as the ratio of the standard deviation  $\sigma$  to the mean  $\mu$ . As a number without dimensions, the coefficient of variation allows for a comparison of populations that have significantly different mean values. This measure is often used to discuss the normal

distribution for positive mean values that have a standard deviation significantly less than the mean. The coefficient of variation, as is also the case with the Theil index, tends to be very sensitive to changes at the top part of the income distribution.

Depending on which measure is used, the degree to which there is inequality will vary. All of these measures belong to groups of relative inequality measures that are not necessarily sensitive to relative changes in income scale, but do imply some *a priori* value judgment about the distribution itself. On the basis of data from the Luxembourg Income Study (LIS) Brigitte Buhmann, Lee Rainwater, Guenther Schmaus, and Timothy Smeeding (1988)) found that when using the Atkinson index, the United States had the highest inequality in disposable income, followed by the Netherlands and Australia. Sweden and Norway, by contrast, had the most equally distributed income. And yet, on the basis of both the Gini coefficient and the coefficient of variation, the Netherlands actually had more equality relative to other countries.

To these four measures may be added yet another: the ratio of say the mean or median income of those at the top of the distribution to the mean or median of those at the bottom of the income distribution. This can be done through a quintile analysis or perhaps even a decile one. The problem with measuring inequality, regardless of which approach is taken, is data. Researchers often rely on official income figures from say a national census. In the United States, for instance, this can be quite problematic because the U.S. Census bureau top-codes the income variable usually at \$1 million, whether it be personal, household or family. This means that in a city like New York City, for instance, where there may be a disproportionate number of those earning

more than \$1 million, the extent to which there is income inequality is understated (Levin-Waldman 2001).

Various researchers have tried to get around this problem by using 10-50-90 percentile levels to analyze the distribution rather than quintile distribution. This avoids the top-coding problem largely because it excludes those over the 90 percentile level. And yet, most of the skew comes from that top 1 percent that has now been eliminated. Even though these techniques will avoid understatement of income inequality, the fact remains that a comparison of the 10-90 percentiles of the income distribution still eliminates the top 10 percent where a disproportionate share of family income happens to be. As Table 1 makes clear, those countries with high Gini coefficients also have significant proportions of wealth and/or income held by the top 10 percent. An example of inequality using this approach for OECD countries during the mid 1980s can be seen in Table 2. Again on the basis of the ratio of the 90 percentile to the 10 percentile, the United States has the highest rate of inequality at 5.94 compared to Finland with the lowest rate of 2.59.

### **Rising Income Inequality**

To talk about income inequality is somewhat problematic because it isn't entirely clear just what we mean by it. What does it mean to say things are unequal in terms of distribution? Income inequality is often viewed as a problem in a market economy, which allocates income on the basis of several factors including education, experience, innate abilities, incentive, and risk. On the contrary, when these factors are considered income is by and large distributed on the basis of desert. More educated individuals, and those possessing greater abilities, are entitled to earn higher

incomes than those who do not. That one is poor, especially in a society where everyone is presumed to enjoy equal opportunity, is ultimately that individual's responsibility. And yet, the capacity to have greater income exists if there is a willingness to obtain the requisite education and training to command it. Although there may be some agreement that a more equitable distribution of income ought to involve a move to greater equality of income and greater equality of opportunity, the prevailing view, at least in the United States is that there is equality of opportunity (Robinson & Dervis 1977).

According to Leslie McCall (2001), the problem is that we tend to look at the income distribution in more absolute terms whereby we compare those at the top of the distribution to the bottom. In reality, there is what she calls "configurations of inequality" in which race, gender, and class intersect in a variety of ways depending on underlying economic conditions in local economies. There is in fact no local economy in which all types of wage inequalities are systematically and simultaneously lower or higher; rather the norm is a complex intersection of various dimensions (p 6). Therefore, inequality needs to be conceptualized as the outcome of both economic restructuring and gender and racial divisions of labor. Inequality will also vary among age cohorts.

Income inequality and poverty are both greater in the U.S. than elsewhere. On the basis of the LIS, Smeeding and Dennis Sullivan (1998) explored differences in economic well-being across cohorts of the population in four modern nations: Canada, United Kingdom, Sweden, and the United States. Using adjusted disposable income (ADPI) as the household's principal measure of well-being, both the U.S. and U.K. were found to have experienced between 1974-94

rapid secular increases in inequality and in relative poverty. Inequality increased less in Canada and Sweden, where poverty levels were also lower. As they observe, overall levels of inequality differ markedly across these nations, and the differences are reflected in their poverty rates.

While the U.S. has had the highest relative poverty rates, Sweden has had the lowest. Younger households (under age 30) in the U.S. are not doing as well as older ones (over 65), but both the young and the old have higher poverty rates than other nations. Those with the highest incomes and the lowest poverty rates in every nation are middle-aged (40-54) families. John Coder, Lee Rainwater and Timothy Smeeding (1989) found that in a comparison of ten modern nations, the most equally disposable income was in Sweden and Norway, and that the highest degree of inequality was in the U.S., Australia, and Canada. Newer countries like the U.S., Australia and Canada tend to have more poverty and fewer people in the middle class than do older nations, and the Scandinavian countries tend to have the least poverty and the largest middle class. But among the elderly, the U.S. while it ranks behind the U.K. and Israel, has the highest poverty rate. It also had a higher fraction of children living in poverty.

Despite a consistent increase in inequality in the U.S. since 1979, the rate of increase has not been constant. The sharpest increase occurred during the early 1980s and was followed by a flattening during the later 1980s. Then during the 1990s income inequality began to re-accelerate (Bernstein & Mishel 1997). The average weekly wages of men, for instance, increased by about 20 percent between 1963 and 1989, but these gains were not spread equally. Wages for the least skilled, as measured by the 10<sup>th</sup> percentile of the wage distribution, fell by

about 5 percent while the wages of the most skilled, as measured by the 90<sup>th</sup> percentile, increased by about 40 percent (Juhn et al 1993). The net result of this divergence was an enormous increase in wage inequality. At the same time, many of the jobs that were created were at the low end of the wage scale. The majority of these low-wage workers in the U.S. have no educational credentials beyond a high-school diploma, and many, including a large number of immigrants, lack even this credential. Eileen Appelbaum et al (2003) refer to these workers as frontline workers. At the same time, a college education and low-wage work are not necessarily mutually exclusive.

### **Sources of Inequality**

The theory of perfectly competitive markets blames rising inequality on structural changes in the economy that have resulted from a mismatch between good paying jobs and the skills available to workers. The main culprit is technological change biased towards those with higher levels of education and skills (Juhn et al 1993). According to this school of thought, the labor market is divided into a primary market where high premiums are placed on skilled workers, and a secondary market where unskilled workers are trapped in the lowest-wage service sector of the economy. The growth in wage inequality between the primary and secondary labor markets has been caused by increasing skills differentials between the two (Katz & Murphy 1992; Katz & Krueger 1992).

Increasingly, greater attention in the literature is paid to institutional factors like wage-setting institutions and social norms. Institutionalists hold rising wage inequality to be due to a shift in public policy and a corresponding decline in labor market institutions like unions and the minimum

wage in the U.S. and wage councils in Britain (Piore 1995; Gordon, 1996; DiNardo & Lemieux, 1997; Fortin & Lemieux 1997; Lee 1997; Machin 1997; Galbraith 1998; Palley 1998; Lemieux 1998; Howell 1999; Wallerstein 1999; Craypo and Cormier 2000). During the late 1970s, the United States began experiencing a sharp ideological shift towards a preference for competitive market outcomes and solutions, and this ideological shift did have direct effects on bargaining in the workplace (Moody 1988).

Examining trends in overall wage inequality in the United States labor market on the basis of data from the Current Population Survey (CPS), David Card and John DiNardo (2002) attributed overall wage inequality in the U.S. at least during the early 1980s to trends in the minimum wage and declining unionization. And in data on worker literacy in OECD countries, David Howell and Friedrich Huebler (2001) found that while there was a positive association between skills differentials and changes in wage inequality, there was also a strong association between labor market institutions and changes in wage inequality. This would suggest a global role for institutions to affect inequality.

Peter Gottschalk (1997) suggests that income inequality increases when the growth of income is greater among those at the top than among those at the bottom, even though bottom incomes have improved in absolute terms. While mean wages grew rapidly during the 1950s and 1960s in the U.S., the dispersion around the growing mean changed very little. But as mean wages grew slowly during the 1970s through the 1990s inequality rapidly increased. So long as those at the bottom of the income distribution gained along with everyone else from secular growth in the mean, it was a

foregone conclusion that poverty rates would be kept down. Moreover, countries with the greatest increases in income inequality were also those that had the most decentralized labor markets. On the contrary, those countries with centralized wage-setting institutions tended to have less income inequality. And as data from the LIS demonstrates, those countries with greater welfare provision and other welfare state institutions that serve to boost the incomes of those at the bottom tend to have lower levels of income inequality (Kenworthy 1999).

### **Social Effects of Inequality**

The most visible effect of income inequality is perhaps poverty, but because poverty isn't something that affects most people we may not easily see the connection between inequality and poverty. And yet, as Mark Rank (2004) argues, poverty, just like income inequality, results from failings in economic and social structures. Much of the current research does establish that individual and family characteristics do have an effect on whether one is likely to experience poverty during one's lifetime. There are particular attributes and characteristics that place individuals at a disadvantage in the labor market.

Kathryn Neckerman (2004) argues that economic inequality by itself would of course be a cause for concern, but its impact is only compounded by its social consequences. There has been a rise in social inequality in many different realms such as family life, education, or civic engagement. Social inequality in these realms only magnifies the burden of rising poverty for the most vulnerable, and thus sustains for a long time the effects of inequality. Neighborhoods with high concentrations of poverty and scarce social resources have

been linked to developmental problems for children living in them. Educational inequality has its most immediate consequences in the labor market with new workers being sorted into good jobs and bad jobs. Then there are perhaps effects of economic inequality on the political system.

Between the 1970s and the 1990s, in the U.S. for example, at a time when social and economic inequality was increasing, there was also a divergence of family patterns across social and economic strata. A well known changing family pattern was a shift in family structure. There was a dramatic increase in single-parent families, and it is this particular shift in family structure that has been identified as a factor exacerbating income inequality. Single parenthood appears to be most common among socially disadvantaged groups, and single parenthood also appears to compound social disadvantage in numerous ways (Martin 2004).

According to David Ellwood and Christopher Jenks (2004) not only has family structure changed from 1900 through the late 1960s, but it has changed very differently depending on parents' education and race. From an economic perspective, the most bewildering feature of family change has been the spread of single-motherhood, which in turn has played a major role in the persistence of poverty. Single-parent families have less income than two-parent families. Although children living with stepparents in adolescence have about the same family income as those living with their biological parents, they are also more likely to drop out of high school or to have a child while still a teenager. In this vein, children living with a stepparent are at least as disadvantaged as children living with a single parent. Mothers' educational attainment rose substantially between 1960-



2000, but during the first score of that period the percentage of mothers who were unmarried also rose sharply throughout the educational distribution.

Traditional economic models treat marriage as a contract from which both husbands and wives expect to reap economic benefits. Structural changes in the economy no doubt affect those expectations. One of the popular explanations for the rise of single-motherhood is that women are less willing to “put up with” the way men treat them (p.52). But in virtually all models and samples, weaker economic performance is associated with reduced or delayed marriage, and they do find some evidence that improving job opportunities for men has somewhat increased marriage and reduced single-parenthood. Theoretical and empirical literature for women, however, has been far more ambiguous about the effects of female labor market opportunities. But improved opportunities for women may have led them to postpone childbearing.

Steven Martin (2004) suggests that delayed family formation may actually be most common among socially advantaged groups, and it may thus confer comparative advantage in numerous ways. Those who delay family formation tend to be better educated and when they do decide to form families, they are in better financial positions to pay for high quality child care and thus greatly reduce their lost career time, which also carries with it a deficit in income. Economically successful women in particular tend to have a smaller wage penalty associated with having children. Although increases in income inequality have not been the sole, or even the primary, cause of declining early adult marriage and marital childbearing, it does nonetheless appear to be an important factor in how

families adapt to new opportunities and constraints.

Rising income inequality may have been affecting social inequality by increasing residential segregation along income lines as well as ethnicity. Recent research on the neighborhood effects of income inequality suggest that neighborhood characteristics such as poverty, crime, and residential turnover influence several interrelated aspects of the neighborhood environment, which in turn affect families and children. Advanced industrial countries have gone through a process of economic restructuring assumed to be strongly associated with the process of globalization. As a consequence of these larger economic trends which have exacerbated income inequality, there has been social polarization, which also varies from country to country (Musterd & Ostendorf 1998; Hammett 1998). Similarly, Andrew Beer and Clive Forster (2002) note that during the 1980s and 1990s, the Australian government embraced international processes of economic change designed to transform the economy and hasten the emergence of the new economic order. As a consequence, income inequality between 1976-1981 increased as income of the poor fell, but after 1981 it increased as the wealthy became wealthier.

These trends have a particular effect on European cities, where spatial segregation tends to be more visible. European cities have generally been experiencing a growing problem of social exclusion, aggravated by spatial segregation, especially concentrated among disadvantaged groups. These disadvantaged include the unemployed, the young and the unskilled. As much as there may be any number of explanations for social exclusion, a common underlying factor is change in economic structure, stemming from global competition and

technological innovation. Changes in economic structure have resulted in both simultaneously an under-representation of unskilled workers—the most rapidly declining group—and an increase in jobs requiring greater skills (Slouten 2000). And yet, in Western societies polarization tends to be mediated by the structure of welfare provision and taxation.

Sako Musterd and Wim Ostendorf (1998) note that in Amsterdam, social exclusion related to the lack of social participation is especially a problem. Social exclusion, then, is one of the most important potential consequences of many of those processes related to social problems. In Sweden disadvantaged people tend to cluster voluntarily or involuntarily, in isolation from mainstream social and economic activities. Addressing income inequality has long been an overarching political goal. The government has also attempted to mix different groups of households in ‘integrated housing’—ideally a mix of households with different demographics, socio-economic and ethnic characteristics. Swedish welfare policy has also been focused on economic resources. An ideological cornerstone of the Swedish welfare state has been equality between households, despite demographic, socio-economic and ethnic characteristics, as well as residential patterns (Borgegard et al 1998).

Mechanisms of segregation fall into four categories: child and family related institutions; social organization and interaction; normative environment; and labor and marriage markets. Child and family-related institutions include schools, child care providers, public libraries, recreational programs and activities, parks, religious institutions, and social service providers. Economic models, of course, suggest that labor and marriage markets are

key elements in the neighborhood effects on families and children. The normative environment includes neighborhood norms that may be a consequence of characteristics of people who live in the neighborhood such as their income level, ethnic background, education, or immigrant experience.

Norms may also be affected by social organization and interaction and also by marriage and labor markets. The central idea in the normative environmental literature is “that the greater the concentration of like-minded people, the stronger the normative climate and the greater the exposure of neighborhood residents to these norms” (Pebley & Sastry 2004:122). The central question may well be whether children who grow up in poor neighborhoods are worse off than other children. And are disparities in children’s welfare by neighborhood poverty level due to differences in their family characteristics, or do neighborhood conditions themselves play a role?

Robert Haveman et al (2004) maintain that as family income inequality increases, those families below the median are further from the social norm than before. Similarly, those at the top of the distribution see a larger gap between themselves and the rest of the population. Many fear that the growth in income disparities among families has had a variety of adverse consequences for both families and communities. The question is how changes in overall income inequality may affect children’s attainment. What is clear is that the growing economic distance between people can reduce common interests and increase social separation.

Families at the bottom of the distribution may end up drifting further from the mainstream, and thus may also experience greater alienation as those with greater resources may come to see them as both more distinct and undeserving. This may

also have consequences for how citizens in turn view the potential role and functions of government. Haveman et. al. argue that when studying the potential effects of growing family income inequality, it is particularly important to consider trends in the level and inequality of inputs in children's attainments, and also the trends in those attainments. Looking at three key indicators of family inputs—parental education, family structure, and family size—they conclude that the situation for children has improved substantially on both parental education and family size but worsened on family structure.

The standard assessment of labor market performance uses employment and earnings over time, but these indicators combine both opportunities in the labor market that are open to people and their choices regarding labor supply and work. While they found evidence that both the level of family resources and the level of attainment of young people have improved over time, they also found limited evidence that better quality of what might be thought of as social capital, i.e. neighborhood quality and school quality, is associated with better outcomes.

Insofar as schools are financed through property taxes, residential patterns along income lines will also result in unequal schools. Schools with well-stocked libraries, for instance, may because of their symbolism attract higher quality teachers and educationally involved parents who can convey the importance of reading to children. Moreover, inequality increases at the margins every time a high-achieving student, a very involved parent, or highly qualified teacher moves from a poor or predominately minority school district into a whiter or richer one (Phillips & Chin 2004).

Inequality in social capital will consequently further exacerbate inequality to

the extent that it affects access. As a consequence of inequality different groups will be socialized differently. Pierre Bourdieu couches the relationship in terms of what he refers to as *habitus*, a pattern of beliefs and behavior based on the experiences of living among certain classes or groups. As Pekka Sulkunen (1982) explains, “the habitus of a group or a class defines a symbolic order within which it conducts its practices—in everyday life as well as in the feast. It provides a common framework within which the members of the group understand their own and each other’s action” p. 108). Based on *habitus*, the styles of working class individuals, particularly in France, are functional, which is the polar opposite from intellectuals—say university professors—and those in highly educated liberal professions.

The importance of this lies in the ability of those in higher socio-economic classes, i.e. those at the top of the income distribution, to reproduce themselves. Consequently, those at the bottom are placed at a comparative disadvantage. Rudolf Richter (2002) notes that in their analysis of French educational institutions, Bourdieu and his colleagues found that elites reproduce themselves through the control they exercise over institutions. An elite interacts with others among the elite, and through this interaction it reproduces. *Habitus* in this sense means that the system reproduces itself, because those with power reproduce themselves and thus separate themselves from the masses.

Deeply rooted *habitus* gives rise to all specific tastes in food, clothing, art, and so on. David Gartman (1991) suggests that a bourgeois taste for freedom is defined in opposition to a working-class taste for necessity. The class structure of society, which no doubt can be perpetuated through

inequality, becomes embodied in the respective *habitus* of each class because that structure determines the exposure of individuals to different material conditions. As Roy Nash (2003) notes, *habitus* means that social structures arise from the process of socialization, which in turn create as prism through which we view the world. Therefore, if as a result of income inequality different groups are socialized differently, the effect will be for the social structure—unequal ones at that—to reproduce themselves.

Rising income inequality has also had an impact on access to healthcare. The strongest evidence appears to show that those with low levels of income have poorer health than those with higher levels. Nevertheless, income inequality per se is not the main factor affecting health status either in the United States or other countries (Mullahy et al 2004). While there is a strong positive relationship between individual income and individual health, there is less evidence of a relationship between aggregate income and aggregate health (Eibner & Evans 2004).

### **Political Effects of Inequality**

Because income inequality affects resources, the central question is to what extent those without may be inclined to opt out of the political system. A bedrock principle of democratic theory is the notion that all individuals as citizens enjoy the same consideration of their preferences and interests. All citizens have the same access to governing institutions, and at the most nominal level, this access finds expression in one-person one vote, equality before the law, and equal rights when it comes to speech, press and assembly. Voter turnout, for instance, is much higher among the wealthy than among the poor. But as Sidney Verba et

al (2004) note, the participation gap between income groups is even higher among those who take a more active role in electoral politics, whether it be working as a volunteer in a campaign or making campaign contributions. Although the affluent are somewhat more likely to give time to a political campaign than the less well-off, they don't systematically give more time to the poor when they do actually get involved in a campaign. Voter turnout in the United States is lower than elsewhere, in part because it is more difficult to register to vote. But among registered voters, turnout in the U.S. is closer to that in European countries. Some have even suggested that this could be remedied through a policy of mandatory voting, which is the case in some European countries (Lijphart 1997).

The most important individual and contextual factors influencing the extent to which one participates in the political system is the socioeconomic (SES) model of participation, which stresses a strong association between political activities and an individual's income, participation, and especially education. And the single most important source of participation inequality is the cumulative effect of educational differences. Richard Freeman (2004) notes that on a world scale the U.S. ranks 138<sup>th</sup> in turnout among countries holding elections. Cross-country comparisons suggest three reasons for low turnout in the bottom parts of the U.S. distribution. First is the weakness of American trade-unions, and unions usually organize low-income workers to vote. Second, the U.S. has a first-past-the-post two party system, which elicits smaller turnover than proportional representative systems of voting in which minority opinion votes so as to have a voice in legislation. And third, the congressional-presidential system elicits smaller turnout than a

parliamentary system. While the fact that Americans have become more educated, work in high-status occupations, and have higher family income than in the past works to raise turnout, the rising proportion of the adult population who cannot legally vote works to lower turnout. Moreover, increased time constraints on people as a result of work and family commitments may also be a factor in reducing voter turnout.

Ian Shapiro (2003) suggests that it is better to think of democracy as a means of managing power relations so as to minimize domination. The challenge, then, is to devise ways to manage power dimensions of human interaction that limit domination while minimizing interference with non-power dimensions. In modern times democratic control suggests an independent activity that is subjugated to democratic constraint. What differentiates government's activities from those of other social actors involved in activities such as responding to market failure, building infrastructure, providing education, insuring banks, and providing welfare, is the specter of legitimate coercive force. If democracy is about structuring power relations so as to limit domination, it then becomes unnecessary to think of questions about citizenship as different from questions about any other superordinate constraints. At the same time, a conception of democracy predicated on reducing domination must also pay attention to the relationship between the political system, i.e. participation, and the distribution of income and wealth. The question of particular concern is whether, and under what conditions, democracy redistributes to the bottom quintile of the population those who are living—or are in danger of living—in poverty.

Shapiro argues that we have become accustomed to the coexistence of democracy

with substantial inequality. And yet Nineteenth century elites initially opposed the expansion of franchise out of fear that the newly enfranchised electorate would exert political pressure to redistribute downward. But there has been no demonstrable relationship between expanding democratic franchise and downward redistribution. Intuitively one might think that the greater the inequality the more likely it is that there will be effective demand for downward redistribution, but the opposite would actually appear to be true. As inequality rises and passes a certain threshold, downward redistribution becomes less likely, which no doubt has something to do with the poor being less likely to participate in the process in the first place. And yet, the more extreme the income inequality, the greater the psychic distance between the haves and the have-nots. Such psychic distance, of course, speaks to the anomie that the less affluent are likely to experience out of a sense that because they are distinctly different, the system will simply be unresponsive.

For many contemporary democratic theorists, however, there cannot be real political equality unless there is a measure of economic equality. Unequal distribution of wealth and income may adversely affect individuals' ability to participate in the democratic process on the same footing as equals. Unequal distribution in wealth and income may result in procedural inequality to the extent that those lacking in wealth and income may not enjoy the same access to political and policy officials as those who possess wealth and income may enjoy. With greater concentration of wealth at the top, those at the top are in a better position to use their wealth toward the attainment of their political and other ideological objectives (Bachrach & Botwinick 1992:4-5). Unequal

distribution in wealth and income, then, results in unequal access. To the extent that this is true, the democratic state cannot possibly be treating its citizens as though they were on an equal footing. Consequently, inequality affects our ability to be free, as unequal distribution will effectively result in some being able to make choices that others cannot. Those with more resources may be better positioned to pursue their goals and objectives, while those with fewer resources may find that their ability to pursue their goals and objectives is limited as a result.

### **Is Income Inequality Really a Problem?**

To the extent that income inequality leads to various social inequalities and distorts the distribution of power in the political realm, it becomes a policy issue that perhaps needs to be addressed. Aside from the obvious material reasons for alleviating abject poverty, it would also appear that steps perhaps need to be taken when there is an obvious threat to the democratic process. Timothy Gaffaney (2000) maintains that a democratic polity operates on the premise that individuals will be politically autonomous—that they indeed will be citizens. A democratic society does not necessarily have to entail economic equality, but it does have to ensure that conditions for participation are available to all individuals, because only when it does so does it guarantee a universal application of citizenship. In fact, the state must guarantee conditions for full citizenship, which might suggest that it has an obligation to pursue those policies that result in a more equitable distribution of income.

Elizabeth Anderson (1999) suggests that equality is about individuals' relations to others. The aim, then, is not to ensure that people necessarily get what they morally

deserve, but to ensure that they are in relations of equality to one another. The point of equality is to in essence ensure that individuals cannot be exploited and oppressed by others. In theory, then, equality prevents one with greater resources from receiving better treatment than the one with less because both are equal in terms of their respective moral worth. According to her conception of "democratic equality," all law-abiding citizens are entitled to effective access to the social conditions required to maintain their freedom, i.e. their ability to make choices. Therefore, democratic equality seeks to abolish socially created oppression. It views equality as a social relationship. Moreover, individuals are regarded as equals when each accepts the obligation to justify his/her actions according to principles acceptable to others.

Democratic equality, then, does not require the elimination of income inequality once all citizens enjoy a sufficient set of freedoms to function as equals in society. Society does not have to compensate for inferior natural endowments, but it does have to ensure that conditions are such that individuals can function as equals. The state merely needs to pursue those policies and establish a condition necessary to function as equal and autonomous citizens. Citizenship requires more than the functioning as a political agent; it requires participation as an equal in civil society. Although democratic equality does not require the elimination of income inequality per se, it does suggest limits. These limits would be the point at which income could be converted into status inequality. Which is to say, that considerable income inequality could be a threat as it might result in status inequality. But at the same time, policies that limit income inequality, albeit they will never eliminate them altogether, serve to preserve the

necessary conditions for the maintenance of equal social relations among people. Societies prizing real political equality ultimately have to be concerned with growing income inequality.

Still, the question remains with regards to how income inequality could be reduced. Many have noted that countries with more generous welfare provisions do tend to have lower levels of inequality. This would suggest that short of eliminating free markets, the remedy, though by no means perfect, is for an active welfare state that serves to compensate for the failures of the market place, of which income inequality might be one example. More specifically, there is reason to believe that institutions such as unions and wage floors also reduce income inequality, especially if the effect is for incomes of those at the bottom of the distribution to increase at a higher rate relative to those at the top of the distribution (Gottschalk 1997). On the punitive side, inequality could be reduced through higher taxes on the wealthy. This, of course, raises its own political problems. In the end, however, it is important to distinguish between policies that reduce inequality and those that reduce the effects of inequality.

What is clear, as the example of China suggests, is that countries undergoing economic reforms are more likely to see a widening of income inequality. And only if they are prepared to invest in education and the social safety net, as well as having redistributive taxation, may inequality be expected to be reduced (Benjamin, Brandt, Giles and Wang 2005). Growth in inequality arises from economic transformation, and the ability to constrain that growth is a function of a government's commitment through policy to alleviate the burdens of that transformation.

## Selected References

- Anderson, Elizabeth S. (1999) "What is the Point of Equality", *Ethics*, Volume 109 (January), pp. 287-337.
- Appelbaum, Eileen, Annette Bernhardt and Richard J. Murnane. (2003) "Low-Wage America: An Overview." in Eileen Appelbaum, Annette Bernhardt and Richard J. Murnane (Editors), *Low-Wage America: How Employers are Reshaping Opportunities in the Workplace*. New York: Russell Sage Foundation.
- Bachrach, Peter and Aryeh Botwinick. (1992) *Power and Empowerment: A Radical Theory of Participatory Democracy*. Philadelphia: Temple University Press.
- Beer, Andrew and Clive Forster. (2002) "Global Restructuring, the Welfare State and Urban Programmes: Federal Policies and Inequality within Australian Cities", *European Planning Studies*, Volume 10, Number 1, pp.7-25
- Benjamin, Dwayne, Loren Brandt, John Giles and Sangui Wang. (2005) "Income Inequality During China's Economic Transition." Prepared for Loren Brandt and Thomas Rawski (Editors), *China's Economic Transition: Origins, Mechanisms, and Consequences*. Cambridge: Cambridge University Press.
- Bernstein, Jared and Lawrence Mishel. (1997) "Has Wage Inequality Stopped Growing?", *Monthly Labor Review*, December, pp.3-16.
- Borgegard, Lars-Erik, Eva Andersson and Susanne Hjort. (1998) "The Divided City?: Socio-Economic Change in Stockholm Metropolitan Area, 1970-94", in Sako Musterd and Wim Ostendorf (Editors), *Urban Segregation and the Welfare State: Inequality and Exclusion in Western Cities*. London & New York: Routledge.

- Buhmann, Brigitte, Lee Rainwater, Guenther Schmaus and Timothy Smeeding. (1988) "Equivalence Scales, Well-Being, Inequality, and Poverty: Sensitivity Estimates Across Ten Countries Using the Luxembourg Income Study Database", *Review of Income and Wealth*, Volume 34 June, pp.115-142.
- Card, David and John E. DiNardo. (2002) "Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles", *NBER Working Paper No. 8769*. New York: National Bureau of Economic Research (February).
- Coder, John, Lee Rainwater and Timothy Smeeding. (1989) "Inequality Among Children and Elderly in Ten Modern Nations: The United States in an International Context", *American Economic Review Papers and Proceedings*, Volume 79, Number 2 (May), pp. 320-324.
- Craypo, Charles and David Cormier. (2000) "Job Restructuring as a Determinant of Wage Inequality and Working-Poor Households", *Journal of Economic Issues*, Volume 34, Number 1, pp. 21-42.
- DiNardo, John and Thomas Lemieux. (1997) "Diverging Male Wage Inequality in the United States and Canada, 1981-1988: Do Institutions Explain the Difference?", *Industrial and Labor Relations Review*, Volume 50, Number 4, pp. 629-650.
- Eibner, Christine E. and William N. Evans. (2004) "The Income-Health Relationship and the Role of Relative Deprivation." in Kathryn Neckerman ed., *Social Inequality*. New York: Russell Sage Foundation.
- Ellwood, David T. and Christopher Jencks. (2004) "The Uneven Spread of Single-Parent Families: What Do We Know? When do We Look for Answers." in Kathryn Neckerman (Editors), *Social Inequality*. New York: Russell Sage Foundation.
- Fortin, Nicole M. and Thomas Lemieux. (1997) "Institutional Changes and Rising Wage Inequality: Is there a Linkage?", *Journal of Economic Perspectives*, Volume 11, Number 2, pp. 75-96.
- Freeman, Richard B. (2004) "What, Me Vote?" in Kathryn Neckerman ed., *Social Inequality*. New York: Russell Sage Foundation.
- Gaffaney, Timothy J. (2000) *Freedom for the Poor: Welfare and the Foundations of Democratic Citizenship*. Boulder, CO: Westview Press
- Galbraith, James K. (1998) *Created Unequal: The Crisis in American Pay*. New York: The Free Press.
- Gartman, David. (1991) "Culture as Class Symbolization or Mass Reification? A Critique of Bourdieu's *Distinction*", *American Journal of Sociology*, Volume 97, Number 2, pp. 421-427.
- Gordon, David M. (1996) *Fat and Mean: The Corporate Squeeze of Working Americans and the Myth of Managerial "Downsizing"*. New York: The Free Press.
- Gottschalk, Peter. (1997) "Inequality, Income Growth, and Mobility: The Basic Facts", *Journal of Economic Perspectives*, Volume 11, Number 2, pp. 21-40.
- Ha Hammett, Chris. (1998) "Social Polarization, Economic Restructuring and Welfare State Regimes." in Sako Musterd and Wim Ostendorf ed. *Urban Segregation and the Welfare State: Inequality and Exclusion in Western Cities*. London & New York: Routledge.
- Haveman, Robert, Gary Sandefeur, Barbara Wolfe and Andrea Voyer. (2004) "Trends in Children's Attainments and Their Determinants as Family Income Inequality Has Increased." in Kathryn Neckerman



- (Editor), *Social Inequality*. New York: Russell Sage Foundation.
- Howell, David and Friedrich Huebler. (2001) "Trends in Earnings Inequality and Unemployment Across the OECD: Labor Market Institutions and Simple Supply and Demand Stories", *Working Paper No. 23*. New York: Center for Economic Policy Analysis, New School University (May).
- Howell, David. (1999) "Theory-Driven Facts and the Growth in Earnings Inequality", *Review of Radical Political Economics*, Volume 31, Number 1, pp. 54-86.
- Jacobs, Lawrence R. and Theda Skocpol. (2005) "American Democracy in an Era of Rising Inequality" in Lawrence R. Jacobs and Theda Skocpol (Editors), *Inequality and American Democracy: What We Know and What We Need to Learn*. New York: Russell Sage Foundation.
- Juhn, Chinhui, Kevin Murphy and Brooks Pierce. (1993) "Wage Inequality and the Rise in Returns to Skills", *Journal of Political Economy*, Volume 101, Number 3, pp. 410-442.
- Kanbur, Ravi and Xiaobo Zhang. (2001) "Fifty Years of Regional Inequality in China: A Journey Through Revolution, Reform and Openness", *CEPR Discussion Paper No 7236*. Ithaca, New York: Department of Applied Economics and Management, Cornell University.
- Katz, Lawrence and Kevin M. Murphy. (1992) "Changes in Relative Wages, 1963-1987: Supply and Demand Factors", *Quarterly Journal of Economics*, Volume 107, pp. 35-79.
- Katz, Lawrence; Kevin M. Murphy and Alan B. Krueger. (1992) "The Effect of the Minimum Wage on the Fast-Food Industry", *Industrial and Labor Relations Review*. 46, 1 (October):6-21.
- Kenworthy, Lane. (1999) "Do Social-Welfare Policies Reduce Poverty? A Cross-National Assessment", *Social Forces*. Volume 77, Number 3, pp.119-39.
- Lee, David S. (1997) "Wage Inequality in the United States During the 1980s: Rising or Falling Minimum Wage", *Quarterly Journal of Economics*, Volume 114, Number 3, pp. 997-1023.
- Lemieux, Thomas. (1998) "Estimating the Effects of Unions on Wage Inequality in a Panel Data Model with Comparative Advantage and Nonrandom Selection", *Journal of Labor Economics*, Volume 16, Number 2, pp. 261-291.
- Levin-Waldman, Oren M. (2001) "Local Labor Markets, Income Inequality, and Institutional Responses: The Case of New York City", *Regional Labor Review*, Volume 4, Fall, pp. 36-46.
- Lijphart, Arend. (1997) "Unequal Participation: Democracy's Unresolved Dilemma", *American Political Science Review*, Volume 91, Number 1, pp. 1-14.
- Machin, Stephen. (1997) "The Decline of Labour Market Institutions and the Rise in Wage Inequality in Britain", *European Economic Review*. Volume 41, pp. 647-657.
- Martin, Steven P. (2004) "Women's Education and Family Training: Outcomes and Trends Associated with Age and Marriage and First Birth." in Kathryn Neckerman (Editor), *Social Inequality*. New York: Russell Sage Foundation
- McCall, Leslie. (2001) *Complex Inequality: Gender, Class and Race in the New Economy*. London & New York: Routledge.
- Moody, Kim. (1988) *An Injury to All: The Decline of American Unionism*. London and New York: Verso.
- Mullahy, John, Stephanie Robert and Barbara Wolfe. (2004) "Health, Income,

- and Inequality.” in Kathryn Neckerman ed., *Social Inequality*. New York: Russell Sage Foundation.
- Musterd, Sako and Wim Ostendorf. (1998) “Segregation, Polarisation and Social Exclusion in Metropolitan Areas”, in Sako Musterd and Wim Ostendorf (Editors), *Urban Segregation and the Welfare State: Inequality and Exclusion in Western Cities*. London & New York: Routledge.
- Nash, Roy. (2003) “Social Explanation and Socialization: On Bourdieu and the Structure, Disposition, Practice Scheme”, *The Sociological Review*, Volume 51, Number 1, pp. 43-62.
- Neckerman, Kathryn. (2004) “Introduction.” in Kathryn Neckerman (Editor), *Social Inequality*. New York: Russell Sage Foundation.
- Palley, Thomas I. (1998) *Plenty of Nothing: The Downsizing of the American Dream and the Case for Structural Keynesianism*. Princeton: Princeton University Press.
- Pebley, Anne R. and Narayan Sastry. (2004) “Neighborhoods, Poverty, and Children’s Well-Being.” in Kathryn Neckerman (Editors), *Social Inequality*. New York: Russell Sage Foundation.
- Phillips, Meredith and Tiffani Chin. (2004) “School Inequality: What Do We Know?” in Kathryn Neckerman (Editor), *Social Inequality*. New York: Russell Sage Foundation.
- Piore, Michael J. (1995) *Beyond Individualism*. Cambridge, MA: Harvard University Press.
- Rank, Mark Robert. (2004) *One Nation, Underprivileged: Why American Poverty Affects Us All*. Oxford & New York: Oxford University Press.
- Richter, Rudolf. (2002) “Pierre Bourdieu and the Penetration of the Social World”, *Innovation*, Volume 15, Number 2, pp. 167-169.
- Robinson, Sherman and Kemal Dervis. (1977) “Income Distribution and Socioeconomic Mobility: A Framework for Analysis and Planning”, *Journal of Development Studies* Volume 14, Number, 4, pp. 347-364.
- Shapiro, Ian. (2003) *The State of Democratic Theory*. Princeton: Princeton University Press.
- Slouten, Paul. (2000) “Strategic Planning, Sustainability, and Urban Regeneration.” in M. Carmona and J. Rosemann (Editor), *Globalization, Urban Form and Governance: Second Industrial Conference ALFA-IBIS Proceedings*. The Netherlands: Delft University Press.
- Smeeding, Timothy M. and Dennis H. Sullivan. (1998) “Generations and the Distribution of Economic Well-Being: A Cross-National View”, *American Economic Review: Papers and Proceedings*, Volume 88, Number 2, May, ppp. 254-258.
- Smeeding, Timothy M., Michael O’Higgins and Lee Rainwater. (1990) (Editors) *Poverty, Inequality and Income Distribution in Comparative Perspective: The Luxembourg Income Study (LIS)*. Washington: The Urban Institute Press.
- Sulkunen, Pekka. 1982. “Society Made Visible—On the Cultural Sociology of Pierre Bourdieu”, *Acta Sociologica*, Volume 25, Number 2, pp. 103-115.
- Verba, Sidney, Kay Lehman Scholzman, and Henry E. Brady. (2004) “Political Equality: What do We Know About It?” in Kathryn Neckerman (Editor), *Social Inequality*. New York: Russell Sage Foundation.
- Xiaolu, Wang. (2006) “Income Inequality in China and its Influencing Factors.” *Research Paper No. 2006/126*. World Institute for Development Economics Research, United Nations University.

Wallerstein, Michael. (1999) "Wage-Setting Institutions and Pay Inequality in Advanced Industrial Societies", *American Journal of Political Science*, Volume 43, Number 3, pp. 649-680.

*Oren M. Levin-Waldman*  
*Metropolitan College of New York*  
*New York City*  
*OLEvin-Waldman@mcny.edu*

## Inflation

*Sergio Rossi*

### **Introduction**

Inflation is a decline in the purchasing power of money. It puts an upward pressure on all prices in the national economy, owing to the lower value of each money unit. As a result, the general price level rises, or does not go down as it should when the purchasing power of a unit of money rises on account of some technical progress leading to an increase in total factor productivity and hence a decrease in production costs per unit of output.

All traditional approaches to inflation stem from the axiom that the loss in the purchasing power of money can be measured referring to index numbers and considering their variation in time. “Economists’ perceptions of inflation rest on measurements of the ‘general price level’ and on rates of change of price indexes” (Gale 1981:2).

In the last 15 years or so, there has been a general reduction of measured inflation rates around the world. In most OECD countries, the rate of inflation is today close to or below 3 per cent (per annum). This was not the case in the past. Beside the hyperinflation episodes hitting Austria (1921–22), Russia (1921–24), Germany (1922–23), Hungary (1923–24), and Poland (1923–24), in which the rate of price increase exceeded 50 per cent per month (see Cagan 1956), and the very high inflation rates measured during World War II (1939–45) in a number of countries (see Friedman 1952), the advanced world has experienced creeping inflation at the end of the Korean war (1953) and up to the end of the 1980s (Table 1). In particular, after 1967, a number of developed countries were hit by

an accelerating rate of inflation, which in 1974–75 reached a peak of 23 per cent in Japan and of 24 per cent in the United Kingdom.

*Table 1.* Inflation Rates, Selected OECD Nations (CPI, averages, per cent per annum)

<i>Years</i>	<i>USA</i>	<i>UK</i>	<i>Germany</i>	<i>France</i>	<i>Japan</i>
1953–66	1.48	3.40	2.23	3.94	3.86
1967–74	5.36	7.50	4.27	6.48	7.51
1975–89	6.39	9.86	3.31	8.30	4.17
1990–2003	2.87	3.19	2.73	1.86	0.68

*Source:* Adapted from: Datastream; Organisation for Economic Co-operation and Development, *Main Economic Indicators*, Paris: OECD, various years. West Germany data before 1990; author’s calculation.

The breakdown of the fixed exchange rate arrangement in 1973 modified this trend. As a matter of fact, participating countries lost the nominal anchor that this regime provided, and had to choose another anchor for monetary policy. This gave rise to several policy rules, the first and most famous of which was put to the fore by Friedman (1968), according to whom the rate of growth of the money supply must be made equal to the rate of growth of real GDP (minus the rate of variation in the velocity of circulation of money, see below), to guarantee price level stability.

Friedman’s rule was applied in the second half of the 1970s and most of the 1980s by a number of advanced countries, following the example of West Germany and Switzerland, which adopted it as soon as 1974 and until the end of the 1990s, with different outcomes in terms of expected and measured inflation (see Bofinger 2001:249–53, Rich 2003). Be that as it may, this rule led to a general reduction of inflation rates, from double- to single-digit figures, in the 1980s. This trend continued in the 1990s, which had been characterised by a change in monetary policy strategies towards the adoption of an explicit inflation target.

The problem of inflation, however, begins as soon as one tries to measure its extent, that is to say, already before considering its causes and strategies put into practice to control it.

### Measuring Inflation

Owing to the different ways of measuring prices and of calculating price levels, there exist several different measures of inflation in both theory and practice. The most commonly used, and monetary-policy targeted, measure of the inflation rate is the percentage rate of change in a country's Consumer Price Index (CPI). Other, less used, measures refer to the GDP deflator, or to the Producer Price Index (PPI), which differ from the CPI in respect of data coverage (goods and prices paid).

#### *Consumer Price Index*

Since the existence of national price statistics, inflation has been grasped by aggregating the prices of those items (real goods and services) that are deemed as representative of consumption habits and patterns in the country considered. The percentage rate of change in the resulting aggregate price, or price level, is then deemed as the expression of the inflation rate suffered by income holders in the national economy as a whole. This approach is therefore based on the so-called microeconomic foundations of macroeconomics.

The first, and up to date most widespread, method used in this respect was proposed by Laspeyres in the second half of the nineteenth century. His now well-known index number is obtained as follows:

$$P_L \equiv \frac{\sum_{i=1}^n p_i^1 q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} \times 100$$

where  $P_L$  is the Laspeyres price index,  $i$  represents the item considered (a commodity or service),  $p_i^1$  is this item's price in period 1,  $p_i^0$  is the same item's price in period 0, and  $q_i^0$  is the quantity of this item bought in period 0. In a nutshell, the Laspeyres price index shows how much it costs for consumers in period 1 (the current period) to buy the same basket of goods that they bought in period 0 (reference period). The difference between the value of  $P_L$  for the (current) period considered and 100 (the price index for the reference period, 0) is then an operational estimation of the current inflation rate (in percentage points).

Despite his seminal attempt at improving the mathematical estimation of the average movement in prices in the national economy, Laspeyres (1871:302, our translation) was, in fact, well aware that "with these figures one cannot appraise the absolute decline in the purchasing power of money." In Laspeyres's view, price index analysis could only roughly estimate the decline in the purchasing power of money (p. 309), owing to the various and restrictive assumptions on which this analysis is grounded. This view has been restated by Edgeworth (1925:380) some fifty years later, when he claimed quite provocatively that "the opinion of representative house-wives would be preferable to the formulae of mathematical statistics" in order to assess the decline in the purchasing power of money.

Following Laspeyres, other index number formulae were put forward by a number of authors, the most famous of which are today those of Paasche (1874), Fisher (1922), and Törnqvist (1936). Unfortunately, these index

number formulae lack a clear macroeconomic meaning (Triplett 1980:569)—although their mathematical sense is absolutely clear and not controversial at all. As a result, they cannot be put to practical use for measuring money's purchasing power and its variation over time. Hence, the Laspeyres method is the approach that has been used to date in order to calculate the inflation rate on the basis of a somehow defined representative basket of consumption items, including a number of imported goods that result from the households' consumption surveys, which official statistical offices carry out periodically.

In this respect, the problem is that changes in the CPI actually do not measure the loss in the purchasing power of money. This loss, in fact, results from a change in the relationship between total domestic output (consumption goods, but also investment goods, and public sector goods) and the stock of national money units that support total domestic demand. By contrast, changes in the CPI are the result of movements in the prices of those items that representative households buy at the surveyed points of sale, which may include commercial imports but exclude exported goods as well as investment and public sector goods produced and consumed within the country (see Rossi 2001:chs 1–2).

As a result, the CPI's variations over time measure the price change of a given basket of consumption items (produced in the country or abroad) purchased by selected households over a given period of time, rather than the inflation rate affecting the numerical relation between the money stock issued in a country and the whole output produced in it. This has led a number of economists and institutions to plead for the establishment of some sort of Cost-Of-

Living Index (COLI), to measure the level of consumer prices in the country (see Boskin et al. 1998).

### *The Cost-Of-Living Index*

“By the expression ‘cost of living’ we mean the monetary value of those consumers’ goods which are *in fact* consumed in the course of a certain period of time by an average family belonging to a given stratum of a population” (Konüs 1939:10). In short, a COLI is the comparison of the minimum cost a household must support to achieve the same standard of living—hence to be on the same indifference curve—across two different sets of prices. This conception is thus based on the existence of equivalent baskets rather than on a unique basket of goods for a given level of consumers’ well-being.

As a matter of fact, the CPI obtains under the very restrictive, unrealistic hypothesis that representative consumers buy the same basket of consumption items as time goes by (same goods, of the same quality, at the same points of sale). By way of contrast, the COLI obtains under the more realistic assumption that any standard of living may be reached by various assortments of goods. To be sure, the cost of keeping a given standard of living might go down while the price of a particular basket of goods might go up, or vice versa (see Gordon 1993).

While the COLI concept is theoretically an improvement of the CPI, it nonetheless offers no better analysis for inflation in practice. A separate COLI could indeed be designed for each and every household on the ground of its actual consumption basket and prices paid (Pollak 1998:69). Mathematical statisticians, political economists as well as national policy makers recognize today that “[t]he whole of the theory of the cost-of-living index relates to

an individual consumer” (Triplett 1975:65). In fact, price level analysis never deals with the case of a single income holder, be it a household or a consumer. “When it comes to dealing with groups of consumers, it is less clear what should be done, even in principle”, in order to assess the percentage change in the computed price index, hence the inflation rate (Fisher & Griliches 1995:230).

### *Some Macroeconomic Problems*

On purely empirical grounds, concern has been raised about the inability of a particular price aggregate, or index, to be representative of different forms of consumers’ behaviour, and hence of the relevant cost of living and its variations over time. For instance, it is today well known that “[t]he substantial differences in expenditure shares suggest that in a period of differences in rates of inflation among different commodities, the inflation rate for different age groups may differ” (Boskin & Hurd 1985:441). Despite heterogeneity in the consumption pattern of different age groups, in fact, periodic indexation of social security benefits, like the retirement compensation to the elderly, is carried out on the basis of the percentage rate of change in the all-items CPI as a rough estimation of the periodic increase in the cost of living of the elderly (Hagemann 1982:495).

Now, even if recently a number of national statistical offices have been setting up a series of CPIs for particular groups of consumers (elderly, young, unemployed, single, single-parent families, and so on), and may even go as far as establishing these particular CPIs on a local basis, the underlying assumption of consumers’ homogeneity within each group of agents is disproved by empirical evidence: observed price differences of consumption bundles

purchased by different subgroups of the population in the same country are in fact smaller between these subgroups than within them (see Michael 1979, Hagemann 1982). In addition, for each subgroup of consumers the representativeness of the data collected might turn out to be highly unstable over time, since tastes and habits can change as time goes by.

A further, essential problem remains when measuring inflation through an index number: there may be some cases where the price level varies without there being inflation, and cases where there can be inflation even if the price level does not vary over time. As the present Governor of the Bank of England pointed out, “changes in indirect taxes or commodity prices often affect the domestic price level, but do not in themselves change the underlying rate of inflation” (King 1997:438).

For instance, an increase in indirect taxes is likely to lead to an increase in the targeted CPI, because the commodities and services subject to higher taxation become really more expensive on the marketplace. This increase in retail prices, and probably in the targeted CPI, however, does not affect the purchasing power of the money stock existing within the country. While the purchasing power of those consumers making use of the taxed goods (for example fuel, tobacco, alcohol) is affected negatively by the decision of the state to rise indirect taxes, the government sector obtains exactly that part of purchasing power handed over by the private sector, so that the relation between money and output in the economy as a whole is exactly what it used to be before indirect taxes were raised.

As a further example of the problematical use of price index analysis to assess inflation, consider the case when firms raise their mark-up, in order for them to increase

their income share (that is, profits). Certainly, the resulting increase in retail prices enables firms to raise profits, *ceteris paribus*, but this does not lead to a loss in money's purchasing power, since firms obtain the fraction of income previously held by some other agents (households): what the latter group of agents loses is transferred to the former group of agents (as a profit), so that the different income distribution between the two groups of agents cannot really affect the purchasing power that the existing money units may exert over domestic output.

Finally, let us point out that the stability of the targeted CPI combined with an increasing productivity of production factors might hide the fact that inflationary pressures operate in the economy. In fact, on the assumption that the mark-up does not vary, technical change leading to a reduction in unitary production costs should lead to a decrease in retail prices. If retail prices do not decrease, and hence the targeted price index is stable over time, this may be ascribed to inflation, which exerts its depressive effect on the purchasing power of money even if this effect cannot be noticed by price index analysis.

### **Explaining Inflation**

Owing to the symptom-based (changes in the price level) definition of inflation, economists have been explaining inflation on the grounds of aggregate demand and aggregate supply. In short, since market prices of produced goods and services are the result of a confrontation between supply and demand, their changes in time are deemed as depicting a disequilibrium between total supply and total demand in the economy.

The most well-known explanations of the measured increases in the price level are

thus the demand-pull and the cost-push views of inflation. Both views have been expanded in many ways in the last three decades, giving rise to several diverse approaches such as the analytical distinction between anticipated and unanticipated inflation, based on the rational expectations hypothesis (Lucas 1972, 1973), or the rigidity-of-wages assumption that New Keynesian authors refer to in order to explain price stickiness and inflation (Fischer 1977; Phelps & Taylor 1977; Taylor 1979, 1980). Further, real business cycle theorists explain price increases by technological and sectoral shocks to aggregate supply (King & Plosser 1984, 1994, Kydland & Prescott 1990). All these developments can indeed be traced back to either demand-side or supply-side analysis. As a matter of fact, they all stem from either the demand-pull or cost-push views of price increases, under which we can subsume them here for a brief critical analysis.

#### *The Demand-Pull View*

The traditional explanation of inflation, in the sense of an increase in the targeted price level over time, relies on demand factors. The most often quoted factors are changes in consumer behaviour, bank credit, and public deficits. In all these cases, adherents to the demand-pull view of inflation argue that "too much money is chasing too few goods." As a result, owing to an excessive demand in the economy with respect to produced goods, the general price level rises over time and this is considered as inflation (Machlup 1960; Brunner et al. 1973; Barro & Grossman 1974).

As concerns consumption, a sudden, sharp increase in consumer expenditures is viewed as inflationary, since it increases total demand for produced output without a simultaneous and identical increase in



aggregate supply. As a result, prices are pushed upwards in order to re-establish numerically equilibrium between demand and supply on the marketplace. This phenomenon has occurred, for instance, in the United States after World War II, when the large amount of forced savings accumulated during that war (which financed government expenditures via purchase of treasury bills by households) was spent once the war was over. As a matter of history, from 1946 to 1948 the measured rate of price increases in the United States was around 25 per cent per annum (see Cagan 1956).

In fact, if prices rise as a result of excess demand, this can shift income distribution in favour of those firms selling the items whose demand has increased, but cannot modify the purchasing power of money (see above). To be sure, the latter has not to be confused with the purchasing power of consumers. Recall that money's purchasing power is established by the relationship between, on the one hand, the total supply of commodities and services produced within the country and, on the other hand, the number of money units available in this same country. If retail prices rise, for the reason pointed out above, this merely means that the distribution of money's purchasing power changes, in favour of (some) firms and to the detriment of (some) consumers (Rossi 2001:40). The former agents now possess a part of the purchasing power that the latter agents owned earlier, so that, for the economy as a whole, this purchasing power is not at all reduced but just distributed differently.

Now, if banks step in and expand credit, in particular to consumers, this may give rise to some inflationary pressure on prices in so far as it creates excess demand on the market for produced goods, or on the assets market

(see Chick 1986; Howells 1996). While this is true and should be avoided to some extent, but not fully since bank credit also helps accelerating output sale, this kind of inflationary pressure is not very troublesome. As a matter of fact, bank credit must be repaid sooner or later by the borrower. An increase in bank deposits as a result of bank credit expansion, therefore, is necessarily linked to an identical reduction of the available amount of bank deposits when the relevant credit is reimbursed to the bank. If we thus consider that all "actions that cause *repayment* of loans cause a reduction in deposits" (Howells 1995:100), and also that "[a]t any particular time, existing loans are being repaid while new loans are being demanded" (p. 90), then we may conclude that, even if bank overdrafts are a cause of temporary inflation, for they alter the money–output relationship, they do not alter it on a permanent basis: an inflationary pressure created by excessive bank credit is bound to be compensated by a deflationary pressure in the periods when bank loans are being repaid, which can be seen as a cyclical phenomenon (see Cencini 1995:63).

The same conclusion can be reached when the central bank expands the available amount of settlement balances in the interbank market to avoid liquidity problems for a bank or for the banking sector: those banks relinquishing securities against central bank money dispose of an illiquid stock of wealth (in the form of financial assets), to transform it into a (liquid) means of payment, in order for them to settle their transactions within the domestic banking system (see Rossi 2005 for elaboration).

As regards the third factor, namely, public sector spending and budget deficits, this also is troublesome when it increases prices in the national economy, owing to

excess demand in respect of available output. Now, again, if the state spends more, this means either that some other agents in the economy spend less (say because of an increase in voluntary savings and/or in taxes), or that the state advances an income—via the domestic banking system or the international financial market—that it will earn in some future period and will then use to reimburse this loan. Be that as it may, in all cases the purchasing power of a money unit over total domestic output is not affected permanently but only temporarily in the worst case, that is, when the state obtains a future income through the credit channel, which it will reimburse in some later period. This also occurs if the government obtains an advance of funds by selling treasury bills to the central bank: if it uses these funds in order to finance its demand for goods and services produced by the private sector economy, this exerts an inflationary pressure on the national economy only in so far as this debt is not paid back. Of course, if the government uses this advance of funds to finance a production of public goods, then the upward pressure on prices would not exist at all, because the increase in the money supply is matched in this case by an identical increase in output.

On the whole, demand-pull theories of the inflation mechanism are based on a symptom-based definition of inflationary pressures, and are therefore to be considered with caution in (monetary) policy making, in order to avoid a policy intervention that may harm, rather than benefit, the economic system under scrutiny.

### *The Cost-Push View*

The alternative view of inflation, with respect to demand-side factors, relies on supply-side mechanisms. This alternative view aims to explain inflation, that is, price

level increases, referring to increases in production costs. The most-often quoted source of supply-led price increases is the wage bill, an assumption that was particularly popular in the 1970s among British economists (Balogh 1970, Jones 1973, Wiles 1973, Hicks 1974), and which has been taken up by a host of post-Keynesian authors since then (Moore 1979:57; Davidson 1988:167; Arestis 1997:101; Smithin 1997:403).

This approach has been developed and is now known as the conflict theory of inflation, argued by authors such as Rowthorn (1977), Skott (1989), Dutt (1992), and Burdekin and Burkett (1996). This theory explains inflation referring to “inconsistent claims on income that emerge from the income distribution struggle between workers and firms” (Palley 1996:182). The claims on income by workers and firms may indeed exceed available output at the aggregate level. The excess of income claims over produced output can be a causal factor of the measured rise in prices on the goods market. Workers try to counteract this upward pressure on the price level by bidding for higher money wages, thus setting forth an inflationary (price–wage) spiral in which each party tries to achieve, or maintain, its targeted income share.

In a nutshell, if labour costs, that is to say, money wages increase more than the increase in physical labour productivity over any given period of time, this may lead to an increase in the price level, if firms pass the higher wage bill on to consumers. Advocates of cost-push inflation conclude therefore – with no further analytical elaboration – that this explains the loss in the purchasing power of money.

In fact, also this explanation of inflation is problematic. In macroeconomics, the

measure of output as a whole results from the payment of production costs, which for the set of firms include wages only (as a matter of fact, intra-firms sales and purchases cancel out). As the Keynes idea of the wage-units foreshadowed (Keynes 1936:ch. 4), the payment of labour costs through the banking system associates money with the produced items in such a way that money and output combine into a unique object, that is to say, money income, whose purchasing power is precisely defined by the newly produced output.

Now, if workers obtain an increase in their money wages per unit of labour time, this has a direct effect on the measure of their output, which increases at the same rate as do money wages, indeed independently of the physical productivity of labour (see Rossi 2001:122–31). To be sure, in economic terms labour productivity cannot be measured without resorting to money wages: without money, output is a non-homogeneous heap of physical items that cannot be measured for the purposes of economic analysis (Keynes 1936:38). Hence, if money wages increase, then as a result the number of money units in the economy as well as the value of produced output increase both by the same amount. The money–output relationship, therefore, cannot be affected by wage increases. To repeat, any wage increase elicits both an increase in the number of money units that finance aggregate demand and an increase in produced output, owing to the fact that the latter is measured in wage-units, which are a monetary magnitude.

Owing to the reduction of union powers at the end of the twentieth century, as a result of deregulation and globalisation, conflict theory has lost much of its appeal as an explanation of inflation. In its stead, today, advocates of a cost-push approach to

inflation prefer to refer to the cost of raw materials (such as crude oil) and the cost of imported goods (like fuel for a great number of countries), what we may call here “global inflation” to indicate, in fact, the propagation of recent increases in the general price level across countries all over the world.

All these cost increases can certainly push up the general price level in the country, since domestic producers must spend more to buy intermediate goods, and local consumers must spend more to buy imported items. (Rising oil prices are a case in point.) In fact, neither an increase in the costs of raw materials and/or of intermediate goods, nor an increase in the costs of imported goods can alter the relation between money and output in the country. If the costs of raw materials go up, then the cost of production may increase for the firms that purchase these materials, but sales increase as well, and identically, for those firms that are selling them. Since intra-firms purchases and sales cancel out, the macroeconomic cost of production comprises the wage bill only. If so, then the total sum of wages that firms pay out to workers is the measure of all produced output, and income, hence the definition of a money–output relation that cannot be affected by price variations in raw materials and/or in intermediate products. Again, it is a matter of distinguishing a loss in the purchasing power of money from a loss in the purchasing power of those agents who buy the goods whose price has increased in time. The former loss is namely the definition of inflation, while the latter loss has not to be mixed up with it.

Similarly, if the prices of imported goods increase for the country’s consumers, perhaps owing also to exchange rate movements, this is likely to affect upwards the CPI as well as the cost of living of

consumers. This increase, however, has not to be assimilated to inflation as we know, because the relationship between money and output in the country is unaffected by what concerns foreign output. To be sure, a domestic income has to be spent to purchase foreign output. This income, however, is just transferred from the buyer to the seller, as in a domestic transaction. Now, in the same way as the internal distribution of income between agents (firms, households, and state) does not alter its total amount, domestic income is not reduced by being partially transferred abroad. The fact that a part of the output produced in the country may be purchased by foreigners is irrelevant to determine the purchasing power of money.

In sum, neither the demand-pull view nor the cost-push view is equipped to absolutely understand the inflationary causes of a rising price level: they both reduce inflation analysis to studying the rate of change in price levels, and neglect (indeed ignore) that inflation may have deeper roots than demand-pull and cost-push reasons (see Rossi 2001: Chs 5–6 for an alternative approach to inflation).

### **Controlling Inflation**

Owing to the symptom-based definition of inflation, monetary policy makers have been controlling the rate of variation of the general price level, or its proxy, the CPI, over time. Their monetary policy strategies have notably been steered by the quantity theory of money ( $MV \equiv PT$ ), which is, in fact, a theory of the price level ( $P$ ), given the so-called velocity of money circulation ( $V$ ): on the assumption that the number of transactions ( $T$ ) is exogenous in respect of nominal magnitudes, the rate of growth of the money stock ( $M$ ) determines the growth rate of the price level in the long run. To

keep the latter stable, or at least under control, as time goes by, monetary authorities must control money supply, either directly, by deciding a targeted rate of growth for some monetary aggregate ( $M0$ ,  $M1$ ,  $M2$ , or  $M3$ ), or at least indirectly, by fixing a target inflation rate measured by variations in the CPI.

### *Monetary Targeting*

The first monetary policy strategy that central banks implemented for a long period of time relied on the transmission mechanism that ties the percentage change in  $M$  to the percentage change in  $P$ . To control the rate of change in prices, hence symptom-based inflation rates, central banks decided and published the target rate of growth of a monetary aggregate, so as to limit inflation by exogenously controlling the money stock.

Historically, in the 1970s and 1980s, many central banks had been targeting the growth rate of a broad monetary aggregate (such as  $M2$  or  $M3$ ), hoping in that way to control the rate of expansion of bank deposits, hence to curb inflation. Yet, in light of the difficulties in steering bank credit (loans create deposits) with base money control, and on account of the poor results of this policy with respect to the actual expansion of the targeted monetary aggregate, in the 1980s several central banks modified their strategy and decided to target the growth rate of base money ( $M0$ ), arguing that  $M0$  is under their full control, so that the mismatch between targeted and actual rates of growth of the money base would not be so important as was indeed the case with some broader monetary aggregate. Assuming that a money multiplier exists and that its estimated value remains stable over time, the targeted, and effective growth rates of the money base were thought to lead to

money supply growth in line with the central bank's policy goals.

Now, although the targeted growth rates of M0 were more closely met in many countries than the targeted growth rates for M2 or M3, empirical evidence showed that inflation rates were not better controlled than with a targeted growth rate defined in respect of a monetary aggregate broader than M0. As a matter of fact, money multipliers are unstable over time owing to the unpredictability of the demand for money as well as to financial innovation. This raised some concern as regards a central bank's policy credibility, and led a number of monetary authorities to change the framework of their policy making, to adopt an inflation target explicitly.

### *Inflation Targeting*

Since the early 1990s, an increasing number of central banks have been targeting a rate of (stable and low) inflation explicitly, following the lead of the Reserve Bank of New Zealand in 1990.

A fully-fledged inflation targeting strategy is based on the following elements (Bernanke and Mishkin 1997, Mishkin and Schmidt-Hebbel 2002):

- 1) No nominal anchor apart from inflation;
- 2) An institutional commitment to price level stability;
- 3) Central bank's financial independence from the government (no bail-out);
- 4) Central bank's instrument independence;
- 5) Central bank's accountability and policy transparency.

According to Svensson (1999), this strategy consists in an explicit inflation target and an implicit employment target, with an implicit relative weight on employment stabilisation. It also represents a

half-way station between a rule-based and a fully discretionary monetary policy. As a matter of fact, it may be seen as a discretionary monetary policy strategy under constraint. Those central banks adopting it are in a position to take advantage of flexibility in policy making in order to control the business cycle in the short run, in so far as their actions do not affect price level stability in the long run (see Bernanke et al. 1999).

Targeting an inflation rate instead of some growth rate of a monetary aggregate increases a central bank's transparency, discipline, and accountability, and improves communication with financial markets and the general public. This is particularly needed in those countries, like emerging market economies, which had a relatively poor economic performance in the past (see Fraga et al. 2004).

When a central bank informs the public on the targeted inflation rate, this much increases its policy transparency, which is crucial for its performance. Indeed, transparency improves the efficiency of private sector decision under uncertainty, provided that the central bank is credible. Credibility can be gained if a central bank's monetary policy target is consistent in time with observed inflation rates, and if the monetary authorities can explain in a credible way any deviations from the target. In fact, transparency induces central banks to be fully responsible of their publicly announced target inflation rates. Their accountability in respect of the general public, the financial markets, as well as the central government leads to more discipline in monetary policy decisions, so as to increase their credibility and also to avoid sanctions by the principal (the government or the parliament).

The benefits of adopting an inflation target explicitly are such that expected and actual inflation rates should coincide more often and more easily. This is so much so that the long and variable lags in transmission mechanisms of monetary policy can be reduced, at least to some extent, in order to improve efficiency.

There remain, however, a number of open questions and problems in inflation targeting. The first question concerns the choice and implications of a point target or a target band. To target a precise inflation rate successfully, the central bank must be able to convince the general public as well as the financial markets that its policy instruments are subject to some volatility and that unpredictable shocks might give rise to target over- or undershooting (see Bernanke et al. 1999:294–5 for elaboration). In this case, the central bank's credibility will not be at issue when the actual inflation rate is different from the target. Over the long run, nevertheless, its credibility still depends on its ability to ensure that the target inflation rate is closely met, so as to be able to strongly affect agents' expectations as regards inflation.

By contrast, targeting a band within which the inflation rate should remain, as time goes by, gives more flexibility to a central bank in its decision making process. It also indicates to agents that inflation is not a phenomenon that depends on a central bank's actions only. Adopting a target band, however, elicits some drawbacks: if observed inflation is outside the band, financial markets and the general public interpret this as the result of wrong monetary policy, hence central bank incompetence, more strongly than if a precisely defined target rate of inflation is not met. As a result, the central bank's credibility is more

negatively affected, which will reduce its possibility to influence agents' expectations.

A further problem in this respect concerns the choice of the band width (Jonas and Mishkin 2003). A large range for the targeted rates of inflation indeed increases the chances to meet the target, but also reduces the central bank's faculty to influence the agents' expectations on inflation, because such a wide band limits the *ex-ante* credibility of a non-inflationary monetary policy, which the central bank may be carrying out effectively. As Mishkin (2000) notes, however, it is more likely that a central bank fails to meet a narrow band for targeted inflation, although such a band might far better anchor the agents' expectations on inflation rates, and also offers more flexibility in monetary policy making with respect to a point target.

The second question concerns the horizon of the inflation target. Owing to the long and variable lag of the transmission mechanism of monetary policy, a too short time horizon for the inflation target may be problematical for a number of reasons (see Bernanke et al. 1999:294–5; Mishkin & Schmidt-Hebbel 2002:198–9; Jonas & Mishkin 2003:27–8).

First, a too short time horizon may imply a control problem for the inflation rate. In light of the long lags in monetary policy, a short-term inflation target increases the probability of target failure, even under optimum policy. Further, a too short time horizon may increase instrument volatility. This is so much so that the target band for inflation is narrow. Indeed, if a central bank changes too often its tools in an attempt to meet target inflation, these tools (say, exchange and/or interest rates) become more volatile, destabilising other variables in the macroeconomy, which increases financial instability. Last, but not least, a too short

time horizon means that the central bank does not sufficiently consider employment and output fluctuation in its loss function: to stabilise the latter variables and meet the inflation target, a central bank cannot have a too short horizon, in order to remain flexible enough to react to unforeseen exogenous shocks efficiently.

A number of solutions are available in this respect. The first is to provide some leeway to the central bank in the case of non-attainment of targeted inflation, say under exceptional circumstances that are to be spelt out clearly, possibly *ex-ante*. The second is to consider a measure of core inflation, which excludes the items whose price fluctuates too much over a year, owing to their seasonality or because of supply-side shocks. The third is to extend the horizon considered for inflation targeting, in order to make it consistent with the time lags in monetary policy.

The third question concerns the numerical value of the inflation target. Owing to several measurement problems, price level stability is not to be confused with a zero inflation rate: using Laspeyres's index number formula to determine the CPI and its variations over time leads to an upward bias in measured inflation rates (see Boskin et al. 1998). Further, a long-run positive rate of inflation has some advantages for a country. First, targeting a positive rate of inflation (2 or 3 per cent) reduces the risks of deflation, which increases unemployment and is thus more harmful than inflation. Secondly, if the central bank targets a positive inflation rate, this means that it considers employment and output issues in its social loss function, at least to some extent. This is likely to increase a central bank's support by the general public, which is important for policy performance.

## Conclusion

As Parkin (1987:836) noted, "the theory of inflation is in a fluid state." The very concept of the price level is indeed problematic for the analysis of inflation: no price index can really be used to assess money's purchasing power and its changes in time, as this article shows.

According to most economists, inflation is characterised by an excessive money stock compared to the total stock of saleable output. However, how is this inflationary gap (to be) measured? Reference to the price level and its variations over time would be correct if, and only if, the money stock and total output were autonomous and (at least in part) independent of one another. This, however, cannot be so, neither in theory nor in practice. Beside the fact that the economic measure of produced output cannot be determined independently of money (owing to the physical heterogeneity of all produced goods), no aggregate price level can be used to measure the two things on which it is based, namely, output and money's value. This is indeed a problem of logic.

As a matter of fact, all current economic analyses measure total output sold using some aggregate price level, and the latter is arrived at by the confrontation on the marketplace between this very same output and the money stock. The argument is clearly circular, and is a further proof of the need to reconsider the causes of inflationary pressures anew in order to manage them effectively.

## Selected References

- Arestis, Philip. (1997) *Money, Pricing, Distribution and Economic Integration*. London and New York: Macmillan and St. Martin's Press.
- Balogh, Thomas. (1970) *Labour and Inflation*. London: Fabian Society.

- Barro, Robert J. and Herschel I. Grossman. (1974) "Suppressed Inflation and the Supply Multiplier", *Review of Economic Studies*, Volume 41, Number 1, pp. 87–104.
- Bernanke, Ben S. and Frederic S. Mishkin. (1997) "Inflation Targeting: A New Framework for Monetary Policy", *Journal of Economic Perspectives*, Volume 11, Number 2, pp. 97–116.
- Bernanke, Ben S., Thomas Laubach, Frederic S. Mishkin and Adam S. Posen. (1999) *Inflation Targeting: Lessons from the International Experience*. Princeton: Princeton University Press.
- Bofinger, Peter. (2001) *Monetary Policy: Goals, Institutions, Strategies, and Instruments*. Oxford: Oxford University Press.
- Boskin, Michael J. and Michael D. Hurd. (1985) "Indexing Social Security Benefits: A Separate Price Index for the Elderly?", *Public Finance Quarterly*, Volume 13, Number 4, pp. 436–49.
- Boskin, Michael J., Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches and Dale W. Jorgenson. (1998) "Consumer Prices, the Consumer Price Index, and the Cost of Living", *Journal of Economic Perspectives*, Volume 12, Number 1, pp. 3–26.
- Brunner, Karl, Michele Fratianni, Jerry L. Jordan, Allan H. Meltzer and Manfred J.M. Neumann. (1973) "Fiscal and Monetary Policies in Moderate Inflation: Case Studies of Three Countries", *Journal of Money, Credit and Banking*, Volume 5, Number 1, pp. 313–53.
- Burdekin, Richard C.K. and Paul Burkett. (1996) *Distributional Conflict and Inflation: Theoretical and Historical Perspectives*. London and New York: Macmillan and St. Martin's Press.
- Cagan, Phillip. (1956) "The Monetary Dynamics of Hyperinflation", in Milton Friedman (Editor), *Studies in the Quantity Theory of Money*. Chicago: University of Chicago Press, pp. 25–117.
- Cencini, Alvaro. (1995) *Monetary Theory, National and International*. London and New York: Routledge.
- Chick, Victoria. (1986) "The Evolution of the Banking System and the Theory of Saving, Investment and Interest", *Économies et Sociétés* ("Série Monnaie et Production", 3), Volume 20, Number 8–9, pp. 111–26.
- Davidson, Paul. (1988) "Endogenous Money, the Production Process, and Inflation Analysis", *Économie Appliquée*, Volume 41, Number 1, pp. 151–69.
- Dutt, Amitava K. (1992) "Conflict Inflation, Distribution, Cyclical Accumulation and Crises", *European Journal of Political Economy*, Volume 8, Number 4, pp. 579–97.
- Edgeworth, Francis Y. (1925) "The Plurality of Index-Numbers", *Economic Journal*, Volume 35, Number 139, pp. 379–88.
- Fischer, Stanley. (1977) "Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule", *Journal of Political Economy*, Volume 85, Number 1, pp. 191–205.
- Fisher, Franklin M. and Zvi Griliches. (1995) "Aggregate Price Indices, New Goods, and Generics", *Quarterly Journal of Economics*, Volume 110, Number 1, pp. 229–44.
- Fisher, Irving. (1922) *The Making of Index Numbers*. Boston: Houghton, Mifflin & Co.
- Fraga, Arminio, Ilan Goldfajn and André Minella. (2004) "Inflation Targeting in Emerging Market Economies", in Mark Gertler and Kenneth Rogoff (eds) *NBER*



- Macroeconomics Annual* 2003, Cambridge, MA: MIT Press, pp. 365–400.
- Friedman, Milton. (1952) “Price, Income, and Monetary Changes in Three Wartime Periods”, *American Economic Review*, Volume 42, Number 2, pp. 612–25.
- Friedman, Milton. (1968) “The Role of Monetary Policy”, *American Economic Review*, Volume 58, Number 1, pp. 1–17.
- Gale, William A. (1981) “Introduction”, in William A. Gale (Editor), *Inflation: Causes, Consequents, and Control*. Cambridge Mass.: Oelgeschlager, Gunn & Hain, pp. 1–11.
- Gordon, Robert J. (1993) “Measuring the Aggregate Price Level: Implications for Economic Performance and Policy”, in Kumiharu Shigehara (Editor), *Price Stabilization in the 1990s: Domestic and International Policy Requirements*. Basingstoke: Macmillan, pp. 233–68.
- Hagemann, Robert P. (1982) “The Variability of Inflation Rates Across Household Types”, *Journal of Money, Credit, and Banking*, Volume 14, Number 4, pp. 494–510.
- Hicks, John R. (1974) *The Crisis in Keynesian Economics*, Oxford: Blackwell.
- Howells, Peter G.A. (1995) “The Demand for Endogenous Money”, *Journal of Post Keynesian Economics*, Volume 18, Number 1, pp. 89–106.
- Howells, Peter G.A. (1996) “Endogenous Money and the ‘State of Trade’”, in Philip Arestis (Editor), *Keynes, Money and the Open Economy: Essays in Honour of Paul Davidson*. Cheltenham and Brookfield: Edward Elgar, Volume I, pp. 105–22.
- Jonas, Jiri and Frederic S. Mishkin. (2003) “Inflation Targeting in Transition Countries: Experience and Prospects”, *National Bureau of Economic Research Working Papers*, Number 9667.
- Jones, Aubrey. (1973) *The New Inflation: The Politics of Prices and Incomes*. London: Deutsch.
- Keynes, John M. (1936/1973) *The General Theory of Employment, Interest and Money*, in *The Collected Writings of John Maynard Keynes*. London and Basingstoke: Macmillan, vol. VII.
- King, Mervyn A. (1997) “The Inflation Target Five Years On”, *Bank of England Quarterly Bulletin*, Volume 37, Number 4, pp. 434–42.
- King, Robert G. and Charles I. Plosser. (1984) “Money, Credit, and Prices in a Real Business Cycle”, *American Economic Review*, Volume 74, Number 3, pp. 363–80.
- King, Robert G. and Charles I. Plosser. (1994) “Real Business Cycles and the Test of the Adelmans”, *Journal of Monetary Economics*, Volume 33, Number 2, pp. 405–38.
- Konüs, Alexander A. (1939) “The Problem of the True Index of the Cost of Living”, *Econometrica*, Volume 7, Number 1, pp. 10–29.
- Kydland, Finn E. and Edward C. Prescott. (1990) “Business Cycles: Real Facts and a Monetary Myth”, *Federal Reserve Bank of Minneapolis Quarterly Review*, Volume 14, Number 2, pp. 3–18.
- Laspeyres, Étienne. (1871) “Die Berechnung einer mittleren Waarenpreissteigerung”, *Jahrbücher für Nationalökonomie und Statistik*, Volume 16, pp. 296–314.
- Lucas, Robert E., Jr. (1972) “Expectations and the Neutrality of Money”, *Journal of Economic Theory*, Volume 4, Number 2, pp. 103–24.
- Lucas, Robert E., Jr. (1973) “Some International Evidence on Output–Inflation Tradeoffs”, *American Economic Review*, Volume 63, Number 3, pp. 326–34.

- Machlup, Fritz. (1960) "Another View of Cost-Push and Demand-Pull Inflation", *Review of Economics and Statistics*, Volume 42, Number 2, pp. 125–39.
- Michael, Robert T. (1979) "Variation Across Households in the Rate of Inflation", *Journal of Money, Credit, and Banking*, Volume 11, Number 1, pp. 32–46.
- Mishkin, Frederic S. (2000) "Inflation Targeting in Emerging-Market Countries", *American Economic Review*, Volume 90, Number 2, pp. 105–109.
- Mishkin, Frederic S. and Karl Schmidt-Hebbel. (2002) "One Decade of Inflation Targeting in the World: What Do We Know and What Do We Need to Know?", in Norman Loayza and Raimundo Soto (Editors), *Inflation Targeting: Design, Performance, Challenges*. Santiago: Central Bank of Chile, pp. 171–219.
- Moore, Basil J. (1979) "The Endogenous Money Stock", *Journal of Post Keynesian Economics*, Volume 2, Number 1, pp. 49–70.
- Paasche, Hermann. (1874) "Ueber die Preisentwicklung der letzten Jahre nach den Hamburger Börsennotirungen", *Jahrbücher für Nationalökonomie und Statistik*, Volume 23, pp. 168–78.
- Palley, Thomas I. (1996) *Post Keynesian Economics: Debt, Distribution and the Macro Economy*. London and New York: Macmillan and St. Martin's Press.
- Parkin, Michael. (1987) "Inflation", in John Eatwell, Murray Milgate and Peter Newman (Editors), *The New Palgrave: A Dictionary of Economics*. London and Basingstoke: Macmillan, vol. II, pp. 832–7.
- Phelps, Edmund S. and John B. Taylor. (1977) "Stabilizing Powers of Monetary Policy under Rational Expectations", *Journal of Political Economy*, Volume 85, Number 1, pp. 163–90.
- Pollak, Robert A. (1998) "The Consumer Price Index: A Research Agenda and Three Proposals", *Journal of Economic Perspectives*, Volume 12, Number 1, pp. 69–78.
- Rich, Georg. (2003) "Swiss Monetary Targeting 1974–1996: The Role of Internal Policy Analysis", *European Central Bank Working Paper*, Number 236.
- Rossi, Sergio. (2001) *Money and Inflation: A New Macroeconomic Analysis*. Cheltenham and Northampton: Edward Elgar.
- Rossi, Sergio. (2005) "Central Banking in a Monetary Theory of Production", in Giuseppe Fontana and Riccardo Realfonzo (Editors), *The Monetary Theory of Production: Tradition and Perspectives*. Basingstoke: Palgrave Macmillan, pp. 139–51.
- Rowthorn, Robert E. (1977) "Conflict, Inflation and Money", *Cambridge Journal of Economics*, Volume 1, Number 3, pp. 215–39.
- Skott, Peter. (1989) *Conflict and Effective Demand in Economic Growth*. Cambridge: Cambridge University Press.
- Smithin, John. (1997) "An Alternative Monetary Model of Inflation and Growth", *Review of Political Economy*, Volume 9, Number 4, pp. 395–409.
- Svensson, Lars E.O. (1999) "Price Level Targeting Versus Inflation Targeting: A Free Lunch?", *Journal of Money, Credit, and Banking*, Volume 31, Number 3, pp. 277–95.
- Taylor, John B. (1979) "Staggered Wage Setting in a Macro Model", *American Economic Review*, Volume 69, Number 2, pp. 108–13.
- Taylor, John B. (1980) "Aggregate Dynamics and Staggered Contracts",

- Journal of Political Economy*, Volume 88, Number 1, pp. 1–23.
- Törnqvist, Leo. (1936) “The Bank of Finland’s Consumption Price Index”, *Bank of Finland Monthly Bulletin*, Volume 16, Number 10, pp. 27–34.
- Triplett, Jack E. (1975) “The Measurement of Inflation: A Survey of Research on the Accuracy of Price Indexes”, in Paul H. Earl (Editor), *Analysis of Inflation*. Lexington: Lexington Books, pp. 19–82.
- Triplett, Jack E. (1980) “Comment to A.S. Blinder’s ‘The Consumer Price Index and the Measurement of Recent Inflation’”, *Brookings Papers on Economic Activity*, Volume 11, Number 2, pp. 567–72.
- Wiles, Peter. (1973) “Cost Inflation and the State of Economic Theory”, *Economic Journal*, Volume 83, Number 330, pp. 377–98.

*Sergio Rossi*  
*Department of Political Economy*  
*University of Fribourg*  
*Switzerland*  
*sergio.rossi@unifr.ch*

## Islamic Banking

*Feisal Kahn*

### **Introduction**

Islamic banking is commonly synonymous with interest-free banking but this is a very narrow and limiting definition. To its adherents, it is financial transactions carried out in strict adherence to Koranic principles that command all Muslims to “do good and forbid evil” (El-Gamal 2000:1). Thus, along with *riba* (usually translated as interest or usury), all practicing Muslims must also avoid financial transactions that involve *gharar* (speculation) and *haram* (religiously forbidden) products. Engaging in trade is encouraged but, ideally, profits must be the result of assuming a commensurate share of the risk involved in the transaction by taking an equity stake in the business venture and must not be earned ‘risk-free’ by making a collateralized loan. “All income should be commensurate with work effort” and “money does not create a surplus value by itself” and so interest income is *haram* in Islam (Presley and Sessions 1994:586). In short, “[i]n Islam, one does not lend to make money, and one does not borrow to finance business” (El-Gamal 2000:33).

By one authoritative estimate, there are some 270 ‘Islamic’ banks in the world, holding about US \$270 billion in assets (Benaissa *et al* 2005:1). Although insignificant compared to conventional banks, e.g., the US financial system's ten largest banks alone hold over US \$4.3 trillion in assets (Mishkin 2007:262), over a decade of 15% average annual growth rates (Benaissa *et al* 2005:1) of Islamic banks demonstrates their importance to, and the growing financial clout of, the world's billion-plus Muslims. The popularity of Islamic Banking and Finance among

Muslims is further illustrated by the creation of the Dow Jones Islamic Indexes in 1999 offering “Islamic compliant investment portfolios,” and the number of Western banks, such as Citibank, ABN Amro, Bank of America, Citibank, HSBC, Standard Chartered, and the Union Bank of Switzerland, that either have Islamic Banking subsidiaries or offer Islamic financial products. By some admittedly optimistic estimates, Islamic banks could account for 50% of all savings in the Muslim world by 2010 (Zaher & Hassan 2001:167).

While over 70 countries have Islamic banks (Warde 2000:1), only Pakistan, Iran and Sudan have a fully interest-free banking system; some others, e.g., Malaysia, Bahrain, and Saudi Arabia, have a hybrid banking system where Islamic banks coexist with conventional ones but are usually dwarfed by them. In Malaysia, which is well-integrated into international money markets, Islamic banks and financial institutions only held an estimated 5% of all financial assets in the 1990s (Warde 2000:127). In Bahrain, rapidly emerging as the Muslim world's money center, its Monetary Agency (now renamed the Central Bank of Bahrain) reported that its banking system (mainly off-shore banking) held assets totaling US \$164 billion in July 2006; of these, Islamic financial institutions' assets totaled only US \$ 10.25 billion. However, this represented a staggering 111% average annual growth rate in Islamic banking assets for the period 1998-2005; in contrast, the non-Islamic banking system average annual asset growth rate was only about 6% (Bahrain Monetary Agency 2006). This underscores the growing importance of Islamic banking for the region and the Muslim world. However, this growth is driven by high oil prices and, probably, would decline if oil prices do.

## What is Islamic Banking and Finance?

Kuran (2004a:ch.2) characterizes Islamic Economics, of which Islamic Banking and Finance is an important part, as having three basic tenets (i) the prohibition of *riba* (whether charging or receiving) in all financial transactions; (ii) wealth redistribution, i.e., mandatory levy of *zakat* (religious alms tax) on all movable wealth; and (iii) adherence to Islamic economic norms that “command good” and “forbid evil.” All five main schools of Islamic jurisprudence or *sharia*, four Sunni (*Hanafi*, *Maliki*, *Shafi’i*, and *Hanbali*) and one Shia (*Ja’afri*) are in substantial agreement on these principles. Islamic banking is then, in theory at least, the concrete manifestation of the *sharia* (Islamic law) with regards to finance and banking.

El-Hawary *et al* (2004:5) defines Islamic Banking and Finance, specifically, as a system that adheres to the following four principles:

- 1) *risk-sharing*: the terms of financial transactions need to reflect a symmetrical risk/return distribution each participant to the transaction may face.
- 2) *materiality*: all financial transaction must have a “material finality,” i.e., should be directly linked to a real underlying economic transaction; thus options and most other derivatives are banned.
- 3) *no exploitation*: neither party to the transaction should be exploited.
- 4) *no financing of sinful activities*: transaction cannot be used to produce goods banned by the Koran (e.g., alcohol, pork products, gambling, etc).

The basis for much of Islamic Banking and Finance lies in the vast number of *hadith* (collection of the sayings of Islam's founder, the Prophet Muhammad; a major source of Islamic legal tradition) collected by Islamic

scholars since the Koran itself does not contain much of direct relevance to economics, finance or banking. For example, the following *hadith* from the *Sahih Bukhari, Volume 3, Book 34, Number 383* (available online at USC-MSA; it contains 7,068 ‘authenticated’ *hadiths* in nine volumes) gives an example of the type of injunction common in them:

Narrated [by] Abu Bakra: Allah's Apostle said, "Don't sell gold for gold unless equal in weight, nor silver for silver unless equal in weight, but you could sell gold for silver or silver for gold as you like."

The Koran and *hadith* strictly enjoins Muslims to be scrupulously fair in their business transactions. Islam, unlike Christianity, has historically endorsed commerce and trade as well as making a profit. Prior to his receiving divine revelation, the Prophet Muhammad was a very successful merchant known for his honesty and scrupulously correct business practices.

Usmani (2002:xii,xvii), a former judge of the Pakistani Supreme Court's *Shariat* Appellate Bench and the *sharia* advisor to several major international Islamic banks, argues in his authoritative work that all Islamic financing has to be “asset backed” because “financing in Islam is always based on non-liquid assets which create real assets and inventories” and “matched with corresponding goods and services."

Thus the current strict interpretation of the *hadith* relating to trade and commercial transactions would ban not only speculative financial transactions such as options and futures but also, for example, hedging by forward sale, interest-rate swaps and any transaction involving items not physically in the possession of the seller (e.g., short sales) (see Venardos 2005 and Usmani 2002 for more details). Warde (2000:140-141) takes a

more nuanced approach to derivatives, etc, but generally considers them to be violating Islamic norms. Similarly, activities such as bill discounting and government debt issue with a fixed coupon rate (the basis of modern Open Market Operations by central banks), inflation indexing, securitized debt obligations, and foreign exchange dealings would also be prohibited unless cumbersome intervening steps reminiscent of the medieval *contractum trinius* are taken. The *contractum trinius* was a series of three separate transactions, all of which were permitted by canonical law, to evade the medieval Catholic Church's ban on usury (El Diwaney 2006, El Gamal 2003).

However, these type of restrictions notwithstanding, early Islamic societies (8-12<sup>th</sup> centuries C.E.) developed sophisticated credit and trade mechanisms that were more advanced (roughly three centuries according to Udovitch 1975:7) than contemporary European practices and in some cases directly influenced the development of European ones (Warde 2000). Early Islamic *sarrafs* (trade financiers) appear to have carried out successful financial intermediation and operated an international payment system using letters of credit and promissory notes (El-Hawary *et al* 2004:6). Udovitch (1975:17) details how *ruq'a* (promissory notes) were used to finance local and short-distance trade solely on the merchant's credit. However, Islamic credit practices did not develop the concept of the use-right of money kept for depository safekeeping with a merchant or banker. Udovitch (1975:19-20) argues that Western bankers "not only kept the goods [money] but also had the right to use them for a variety of commercial purposes" and for so doing paid a depository "premium" (i.e., interest) but this restriction on use kept Islamic "credit and banking activities within

rather narrow confines" and contributed to the general ossification of Islamic financial and commercial thought and practices. The remarkable degree of this ossification is shown by the fact that "[i]n eighteenth century Cairo, credit practices hardly differed from those of the tenth century" (Kuran 2004b:71).

However, this view that Islamic commercial and financial practices were the cause of Islamic economic stagnation is not shared by all analysts. Khan (2003:342), although actually arguing for a reinterpretation of classical Islamic law (*fiqh*), states flatly that "[i]t is factually incorrect to argue that classical *fiqh* has been the source of Islamic economic backwardness for great Islamic empires flourished under classical *fiqh*." While factually correct, this view ignores the fact that practices that were highly innovative and efficient under one set of circumstances may no longer be so when circumstances change.

To its advocates, Islamic banking and finance offers a morally superior and economically more efficient alternative to conventional interest-based banking systems. There are essentially two main justifications given for preferring an Islamic economic and financial system to conventional ones. The first may be termed the 'economic efficiency argument' which often relies on work by Western economists (e.g., Milton Friedman's 1969 theoretical argument for a zero nominal interest rate economy) to argue that an equity-finance based economy or an interest-free economic system is one that maximizes overall economic efficiency and societal welfare. The second argument is simply that the banning of *riba* in the Koran (see e.g., Koran 3:130 or 2:275-279) is all the justification necessary. Both views hold that equity

finance is preferable to interest-based finance because if the venture fails, the borrower does not bear the entire cost alone or lose the collateral. If the venture succeeds, the financial investor receives a larger return than a predetermined interest rate would allow. Furthermore, equity finance would make more financial resources available to small businesses, make it more difficult to accrue 'unearned' income and generally promote an ethos of equity and justice in society.

Chapra (2000:33), after a brief survey of the literature, argues that "an economy that relies less on credit and more on equity [financing of businesses] may be superior in its overall performance to the one that relies substantially on credit." El-Gamal (1998:10), after a survey of the economics literature and though very critical of current Islamic banking practices, concludes that an Islamic financial system is theoretically superior to a conventional one because it is (i) more stable, (ii) has more efficient resource allocation, (iii) reduces 'effort aversion' (iv) would finance a broader array of business projects, and (v) adhere to God's commandments. Similar arguments are advanced by other advocates of Islamic Banking and Finance such as Usmani (2002) and Warde (2000).

### **Islamic Banking Practices**

Islamic banking and financial practices can be divided into two main categories: Profit and Loss Sharing (PLS) and Non-Profit and Loss Sharing (Non-PLS) (Sundarajan and Errico 2000:20). That is, Islamic finance is either participatory (an equity stake is taken by the financial institution) or non-participatory (an equity stake is not taken by the financial institution). The two main participatory forms of Islamic finance are *mudaraba*, where a 'sleeping' partner

contributes capital and another expertise/knowledge, and *musharaka*, where there is equity participation. If non-participatory, the most commonly used method is either a *murabaha*, a 'markup' or cost-plus sale, or *ijara*, a lease; *bai salam/istisna*, deferred delivery, *bai muajjal*, deferred payment, *jo'alah*, service fee, and *qard-e-hasana*, beneficence loan, are examples of other possible NonPLS finance forms. Box 1 summarizes these major Islamic financing modes.

While there are many other forms possible, these constitute the majority of Islamic financing, with *murabaha* and *ijara* clearly dominating all others as "[globally] profit and loss sharing transactions... only account for about five per cent of the operations of Islamic financial institutions (Warde 2000:240)." El-Hawary *et al* (2004:15) concur as "[i]n a typical Islamic bank, [nonPLS forms] dominate the assets portfolio and can exceed 80%." Further evidence comes from an analysis of the Saudi-based Islamic Development Bank (IDB), which is perhaps the single most important funding source for research on Islamic Economics and Finance. An analysis of its asset portfolio for 1976-2004 showed that 91% of its financial transactions were non-PLS.

A typical *murabaha* transaction would be structured as follows. A firm in need of upgrading its existing machinery would approach an Islamic bank to purchase the machinery on its behalf, with a concomitant agreement that the firm would purchase the machinery, on a marked-up basis, from the bank. If the bank's purchase price is \$100,000, it might resell the machinery to the firm for \$110,000, payable in 12 equal monthly installments. The bank retains ownership until the last installment is paid and the bank's position is fully secured; this

is accepted as *sharia*-compliant by many authorities (e.g., Usmani 2002:52-54 and Warde 2000:133). A *murabaha* is Islamically permissible since the bank has had actual ownership of the machinery and so has borne some risk; a simple loan for \$100,000 at 10% interest, secured by the machinery, is not. The obvious similarities to traditional interest-based debt finance are clear and El-Gamal concludes in his study of Islamic Finance that "the bulk of Islamic financial practices [worldwide]... would easily be classified by any MBA student as interest-based debt finance" (El-Gamal 2003:108).

Thus *murabaha*, *ijara* and other NonPLS forms are viewed by most *ulema* as weakly Islamic at best. The general consensus among *ulema* is that participatory finance is the desirable alternative and non-participatory finance is acceptable only as an interim measure or for situations where participatory finance is clearly unsuitable, such as very small or personal consumption loans (Usmani 2002:42; see also Kuran 2004; Sundarajan and Errico 2002, and especially Warde 2000; however El Gamal 2000:15 argues that there is no Koranic justification for this preference).

Some critics (e.g., Kuran 2004a) of contemporary Islamic banking practices go further and argue that *murabaha* is merely legalistic sophistry, an Islamic version of the old medieval *contractum trinius*. It was in keeping with this view of *murabaha* as simply 'disguised interest' that the Pakistani courts ruled against the existing system of Islamic banking in 1991 and 1999 and mandated a conversion to full equity participation by the banking system; this was overturned on final appeal in 2002 (Khan 2006).

El Hawary *et al* (2004:16-17) also detail another important divergence between

Islamic banking theory and actual practice. Although ostensibly accepting deposits on a PLS basis since no preset 'interest' rate is permissible, no Islamic bank has ever written-down the value of its depositor's accounts when it has written-down the value of its non-performing assets. In other words, bad debts are not translated into "losses" for depositors. On the contrary, Islamic banks have declared market-competitive returns for depositors even when running into financial difficulties themselves.

Many central banks also implicitly guarantee Islamic bank deposits in a manner not consonant with Islamic risk-sharing principles. For example, a run on Dubai Islamic Bank in 1998, following reports of a massive corruption scandal, resulted in the withdrawal of US \$138 million, 7% of its total deposits, in one day; such a large withdrawal would have caused the Bank's collapse if the Dubai Central Bank had not stepped in as the "lender of last resort" and effectively guaranteed all deposits (Warde 2000:155-156).

Islamic bankers' preference for nonPLS financing is an eminently rational response to the information asymmetry problems inherent in all financial transactions. All such transactions involve some degree of information asymmetry: here the borrower usually has better information than the lender/financier. Consequently, there is a risk of two types of problems. The *ex ante* problem is adverse selection when loans/investments are made to/in a poor credit risk. The *ex post* problem is moral hazard when loaned/invested funds are misused and/or utilized in inappropriate ways.

Naturally the degree of information asymmetry, and the resulting adverse selection and moral hazard problems, depends upon the amount and quality of



information flows between lender or investor and borrower. Institutions such as credit rating bureaus improve the information flow and reduce the likelihood of adverse selection, while civil, criminal and possibly societal sanctions reduce the moral hazard dangers. However, in the absence of institutions to mitigate the information asymmetry issue, fully collateralized debt-finance would be the financing mode preferred by banks to minimize their risk.

Warde (2000:155-157) confirms this serious problem when he details a long list of “Islamic Moral Hazard” problems. These ranged from depositors demanding higher returns but guaranteed principal amounts, to borrowers challenging the permissibility of financial penalties and late fees, to Pakistani borrowers, in a high inflation environment, cheerfully repaying only the principal of the loan on the grounds that interest is forbidden.

These issues are not confined solely to Muslim countries. In even advanced financial systems, the information asymmetry problems are so severe that most financial transactions are debt-based and not equity-based (see Mishkin 2007, especially Ch. 8). For example, the US market for venture capital and investment, among the largest and most sophisticated in the world, was US \$59.9 billion in 2004 (*Euromonitor International* 2005). In contrast, US bank loans as a source of corporate financing are worth several trillion dollars and most debt is collateralized (Mishkin 2007:182-184).

Therefore, even when Islamic banks do take a direct equity stake, it is sometimes no different from a comparable, interest-based debt-finance transaction. For example, one common way of structuring an Islamic mortgage is as a 'declining *musharika*.' In this arrangement, the financier purchases the

house on behalf of the eventual owner (with or without a down payment) and then rents it out to her. The monthly payment contains two distinct components: ‘rental’ for the financier’s ‘share’ of the house and a ‘buyout’ component for its ‘purchase’ (El Gamal 2000:16). The contract ends when the entire financier portion of the house has been purchased. As El-Gamal (2000:16) notes, the entire process “look[s] very much like a conventional mortgage schedule [and] it is trivial to calculate the equivalent interest rate which would make the conventional mortgage payments identical with the diminishing partnership payments.”

In theory, the ‘rent’ charged should be based on comparable homes in the area but, in practice, the rental varies according to the prevailing mortgage interest rate. Preliminary indications are that the implicit mortgage rate charged Islamic borrowers in the US is at least 25 basis points higher than that of comparable conventional mortgages (Healy 2005), an extra \$6,271 over the life of a 30 year, \$150,000 mortgage. However, Islamic mortgages often entail additional costs for the buyer so total costs are often far higher than a conventional mortgage's (Healy 2005).

Rammal (2004) gives a detailed example of how such a diminishing *musharika* would work. A financier purchases a home worth \$150,000; the financier’s share is \$120,000 (80%) and the homeowner contributes \$30,000 (20%). A fair rental price for the home is \$1,000 and so the initial imputed rent accruing to the financier is 80% of \$1,000 or \$800; a monthly ‘extra payment’ is made for purchasing the financier’s share of the house. Table I (Rammal 2004; ignoring insurance, taxes, etc) gives the payment schedule for a 15 year (180 month) purchase plan:

**Table I: Example of Payment Schedule for Home Loan Under *Musharaka***

Month	Rent \$	Extra Payment \$	Total Fixed Payments \$	Bank's Ownership \$
Opening				120,000
1	800	347	1147	119,653
2	798	349	1147	119,304
...	...	...	...	...
176	37	1110	1147	4,439
177	30	1117	1147	3,322
178	22	1125	1147	2,197
179	15	1132	1147	1,065
180	7	1065	1072	0

Contrast this with Table II, a conventional amortization schedule for a 15 year, \$120,000 mortgage at 8% interest (ignoring all taxes and insurance costs). As is readily apparent, the differences between the two are trivial. Thus the differences between actual Islamic banking practices and conventional debt-finance with an interest rate charge are often minimal at best.

**Table 2: Amortization Table for Conventional US \$120,000, 15-year mortgage at 8% p.a.**

Month	Interest Paid \$	Principal Paid \$	Mortgage Payment \$	Balance \$
Open				120,000
1	800	347	1147	119,653
2	798	349	1147	119,304
...	...	...	...	...
176	37	1109	1147	4,439
177	30	1117	1147	3,395
178	23	1124	1147	2,271
179	15	1132	1147	1,139
180	8	1139	1147	0

## Islamic Capital Markets

A *sukuk* is a generic term for an Islamic financial certificate, an 'Islamic bond.' *Sukuks* are corporate or public debt instruments, usually with a variable rate of return tied to a commonly accepted market indicator rate (EURIBOR or LIBOR) and must have an underlying 'real' transaction basis to be *sharia*-compliant. The Accounting and Auditing Organization of Islamic Financial Institutions has issued *sharia*-compliant standards for the issuance of 14 different types of *sukuks* (Glasser 2005). The typical *sukuk* is a 'floater' bond

type of Asset Backed Security with some restrictions as to what it may be used to fund (no activity forbidden by Islam).

The most common type of Islamic bond is the *sukuk al ijara*, essentially a lease-revenue bond (Glasser 2005). Here the borrower's tangible asset is 'sold' to the financiers and then 'leased' back to the borrowers, who make regular payments to the financiers from the income stream generated by the asset; this tangibility/income-stream is what makes the transaction 'Islamic.' General revenue bonds issued by governments or corporations are not Islamically permissible since they are non-Asset Backed Securities (there is no direct connection between the security and the income stream accruing to the financiers) backed by nothing other than faith in the government's/corporation's credibility.

Islamic capital markets have been relatively underdeveloped until very recently but, while growing extremely rapidly, are still insignificant compared to regular bond issues. Gassner (2005) estimated the entire 2004 *sukuk* volume to be under US \$7 billion; Hamsawi (2005) estimated the 2005 market size as US \$9 billion, almost a 30% annual growth rate. A January 2006 US \$3.5 billion offering by Dubai Ports Corporation (the largest *sukuk* offering to-date) was 325% oversubscribed and news reports indicate that several Persian Gulf countries are considering massive (US \$50 to 65 billion) petroleum projects financed by a *sukuk* issue because of the extremely attractive (for the borrower) rates resulting from the high demand for them as almost all *sukuk* "offerings are hugely oversubscribed" (AMEinfo 2006). This popularity has led to some offerings being listed on the London Stock Exchange in 2006 and assigned credit ratings by Standard and Poors (AMEinfo 2006), as well as the creation of the Dow

Jones Citigroup Sukuk Index to track *sukuk* issues and serve as a benchmark rate for “Shariah-compliant fixed income investments” (Parker 2006).

However, Islamic capital markets cannot develop fully if there is no deep secondary market for Islamic securities. *Ijara al sukuk* issued by the Bahrain Monetary Agency, the old name of the Central Bank of Bahrain, are apparently the only Islamic bonds for which there exists a substantial secondary market, on the Bahrain Stock Exchange (El Hawary 2004:24). Presumably this is because of the activities of the Bahraini government’s Islamic Liquidity Centre, created in 2002, which attempts to provide market liquidity for Islamic financial instruments by helping to create a secondary market for them (see Bahrain Monetary Agency website [www.bma.gov.bh](http://www.bma.gov.bh) for more details). Hamsawi (2005) notes that while secondary markets do exist for some other *sukuk* issues, trading is extremely limited.

The lack of *shariah*-compliant government-issued financial securities has led some central banks to modify existing rules for Islamic banks. For example, the State Bank of Pakistan has mandated only an 11% liquidity requirement for Islamic banks (versus 15% for conventional banks) since they can only meet the mandated liquidity requirements through non-interest earning cash reserves rather than government paper (WB-IMF 2004:16).

Secondary market development is further constrained by the fact that not all Islamic scholars are in agreement about what constitutes a *sharia*-compliant financial instrument. For example, Malaysian *ulema*, adherents of the relatively more progressive *Shafi’i* school, have *sharia*-certified securitizing Islamic Private Debt Securities (IPDS), i.e. corporate accounts receivable. However, scholars in Pakistan, India and the

cash-flush Persian Gulf have held that IPDS are not *sharia*-compliant since the underlying transaction is a sale of debt and not a tangible asset. This is, according to their interpretation of the *hadith*, *riba*. Therefore, the largest national Islamic corporate debt market, approximately US \$24.5 billion in beginning-2005, is not ‘Islamic’ enough for investors in most of the Muslim world. (Malaysian details from Parker 2006.)

There does apparently exist a potentially very large and hitherto virtually untapped source of Islamic funds that could be made available relatively cheaply to all borrowers if the *sukuk* market continues to grow and develop. These funds are available to non-Muslim borrowers as well: in 2005 the German state of Saxony-Anhalt issued a five year, 100 million Euro *sukuk al ijara* at an extremely favorable rate (Gassner 2005). There are many Muslims who would prefer to invest their savings in more secure Islamic *sukuk* rather than in gold or real estate, the traditional venue for Islamic savings for those whose religious views prohibit their purchasing conventional fixed-income securities and whose risk tolerance precludes their investing in equities.

### **Regulating Islamic Banking**

There is no one definitive source of either financial or religious regulation for Islamic Financial Institutions (IFIs). While national central banks have regulations for IFIs in their countries, there are some international standard-setting agencies as well.

In 1991, Bahrain, home to two-thirds of the Persian Gulf’s IFIs, established the Accounting and Auditing Organisation for Islamic Financial Institutions (AAOIFI) which consults with the London-based International Accounting Standards Board (IASB) to ensure that Bahrain-based IFIs use

internationally acceptable accounting standards; for e.g., the AAOIFI's Capital Adequacy Ratio for Bahrain-based IFIs is a modification of the Basel Committee standard, the standard for bank capital adequacy requirements worldwide (El Hawary *et al* 2004:28). Bahrain is also home to the International Islamic Rating Agency which provides *sharia*-compliance certification and credit ratings for IFIs.

For non-Persian Gulf IFIs, the Malaysia-based Islamic Financial Services Board (IFSB) was created in 2002 by the central banks of several Muslim countries to establish common "corporate governance, transparency and disclosure" standards for IFIs (El Hawary 2004:28), and to attempt a harmonization of regulatory standards for IFIs. The current IFSB membership includes the World Bank, the International Monetary Fund and the Bank for International Settlements (the central bankers' clearing house, and cooperation and coordination mechanism) as well as some 91 other central banks and commercial IFIs. The IFSB issues Capital Adequacy, Risk Management and Corporate Governance standards for member IFIs (see [www.ifsb.org/](http://www.ifsb.org/) for more details). The AAOIFI and the IFSB appear to coordinate policies quite closely and there is substantial overlap between their membership. The major differences appear to be that the IFSB takes a more progressive view of what is Islamically-permissible than does the AAOIFI (see, e.g., the earlier discussion on Islamic Private Debt Securities in Malaysia).

However, this coordination between conventional (Western, interest based) central banks and global financial institutions and IFIs has not been without criticism from Islamic scholars. There is apparently a substantial number of *ulema* who hold that IFIs should not be subject to

any regulatory or government agency because these agencies do not recognize the religious imperatives that regulate Islamic banks and treat Islamic financial products as 'loans' rather than equity investments (El Hawary *et al* 2004:29; Warde 2000:197). If these *ulema* should gain strength, it would be very difficult for IFIs to operate in countries unwilling to acknowledge their 'special' status.

El Hawary *et al* (2004:35) also express concern that the majority of IFIs are located in countries where "the quality and transparency of accounting and auditing" is insufficient to give an accurate measure of the risk exposure of IFIs. This makes it difficult to enact appropriate regulation since, even in advanced economies with presumably better accounting and auditing standards, financial crises such as the US Savings and Loan Crisis of the 1980s underscore the importance of avoiding financial mistakes. As the long list of "Islamic moral hazard" detailed by Warde (2000), combined with inadequate accounting, auditing and internal controls, has given rise to periodic scandals which have rocked Islamic banking circles.

The most prominent of these scandals were the 1991 collapse of the Bank of Credit and Commerce International (aka 'Bank of Crooks and Conmen International; while not officially an IFI, it had a very large Islamic Banking Unit), the takeover and recapitalization of Dubai Islamic Bank by the Dubai Government in 1998 after a massive fraud/embezzlement scandal, the takeover by the Egyptian central bank of the International Islamic Bank for Investment and Development, the takeover by the South African central bank of the country's Islamic Bank Limited, and the bailout by the Kuwaiti government of the Kuwait Finance House (see Warde 2000 for details on these).

To what extent these are normal ‘growing pains’ of Islamic banking and to what extent these are issues endemic to IFIs remains to be seen.

### **Rationale for Islamic Banking?**

In 1964, the Pakistani government appointed the Islamic Modernist Prof. Fazlur Rahman (1919-1988; the late Harold H. Swift Distinguished Service Professor of Islamic Thought at the University of Chicago) the first head of the newly-created Central Institute of Islamic Research (CIIR). Rahman (1964) argued that *riba* (literally ‘increase’ in Arabic) in the Koran refers not to interest or even to usury but to a particularly repugnant pre-Islamic Arab custom in which a borrower would be asked, upon his debt’s maturity, “Will you pay up or will you increase?” (Rahman 1964, quoting Imam Malik, c.715–796 CE, founder of the Maliki school of *sharia*). The ‘increase’ would be a doubling of his debt. Failing to make repayment of the doubled debt, he would be given another extension of his debt but at the cost of another ‘increase’ (Rahman 1964:6). This, Rahman argues, is the explanation of the Koranic verse (from the Pickthall translation; the Abdel Haleem translation is essentially identical): “O ye who believe! Devour not usury, doubling and quadrupling (the sum lent). Observe your duty to Allah, that ye may be successful.” (Koran 3:130)

A defaulting debtor and his goods would be auctioned off to satisfy his creditors. Thus Rahman concludes that the Koran bans extreme usury, not interest and the prohibition applies really only to personal consumption loans. El Gamal (2003) has a modern exposition of this argument from Muhammad Sayyid Tantawy, the Grand Imam of Egypt’s Al-Azhar Mosque and the

Grand Sheikh of Al-Azhar University, a leading Sunni Muslim *madrasah*.

This interpretation of the Koranic ban on *riba* is consistent with Posner’s (1980) view of the economics of “primitive economies” and Glaeser and Scheinkman’s (1998) economic analysis of interest rate restrictions. The basic argument is that a ban on interest or usury is a risk-reduction mechanism that arises when there is a missing insurance market and when the majority of the population lives near the subsistence level.

Suppose a person borrows at 10% interest to finance a commercial venture and puts her home up as collateral for the loan. If the venture fails, she has to repay the loan plus accrued interest or lose her home. If she does not have enough other assets to repay the loan, she will lose her home, which would put her and her family at risk. If she could have purchased insurance to guard against the venture’s failure, then her home, wealth and family are safe from the normal risk inherent in any business venture. However, premodern societies lack both comprehensive insurance markets and a social safety net to look after their people; hence the ban on interest or usury.

In modern development economics, the new agrarian economics school has argued that peasant risk-aversion leads them to accept economically inefficient sharecropping contracts (that often stipulate a 50:50 or similar split of the harvest) rather than a set lease rate that would accrue all benefits above the rental rate to the farmer (Todaro 2006:446-449). It appears that peasants are pursuing a maximin strategy. They are not maximizing the potential benefits (all returns above the lease rate accrue to them) but trying to minimize the potential loss (in case of a bad crop year,

they do not owe more to the landowner than a portion of what is actually produced).

Within the Islamic context, the argument is exactly the same. An interest bearing loan used to finance a venture that fails has to be paid back with accrued interest or the collateral might be lost. However a *musharaka* agreement to finance a failed venture results only in the loss of the invested assets and nothing further is owed to the financier. If there are no other social safety nets and no insurance markets, then the Islamic ban on *riba* is a societal survival mechanism and a rational societal response to a missing insurance market problem.

Thus current Islamic banking is a modern restatement of widespread premodern societal practices designed specifically to avoid the worst potential outcomes in case of severe financial setbacks. This is why a ban on interest or usury was found in virtually all ancient societies. The earliest known reference to usury is c. 2,000-1,400 BCE from the ancient Vedic texts and the earliest negative comments are also from Hindu texts dating to c. 600-400 BCE (Visser & McIntosh 1998). Usury was condemned and prohibited in ancient Greek, Roman and Jewish societies well before the advent of Islam, and was a mainstay of Catholic as well as early Protestant doctrine (Visser & McIntosh 1998; Glaeser and Scheinkman 1998). In its day, it was an effective and much needed social protection mechanism but one that is perhaps not as relevant today since the underlying socio-economic circumstances have changed.

However, barring a reinterpretation of *sharia* to allow a more progressive view of the ban on *riba* and greater latitude in financial transactions, Islamic banking will continue as it currently is: providing a fairly wide range of financial products and services that closely mimic conventional

ones. The major difference between it and conventional banking and finance will continue to be the additional layer of transaction costs needed for *sharia* certification.

## **Glossary of Islamic Financing Modes**

### **1. Forms of Participatory/PLS Islamic Finance**

**Mudaraba:** ‘Trustee finance contract’ or passive partnership; one party provides funds while the other provides expertise and management; profits accrued are shared on a pre-agreed basis; losses are borne only by capital provider.

**Musharaka:** ‘Equity participation contract’ providing for profit/loss sharing in the joint business; bank provides a portion of the total funds and all partners may participate in management; profits are distributed in pre-agreed ratios but losses are borne strictly in proportion to respective capital contributions. A ‘declining *Musharaka*’ may be used for installment purchase of, e.g., a house with the buyer paying *pro rata* ‘rent’ to the financier.

**Sukuk:** Islamic bonds, generally paying a ‘LIBOR + X%’ rate but may have a fixed rate if backed by *ijara* or *murabaha* transactions; must be asset-backed and Shariah compliant; sometimes also referred to as *Musharaka Term Finance Certificates*, with a nominal equity component.

**Direct Equity Investment:** e.g., purchasing shares on the open market, etc.

### **2. Forms of Non-Participatory/Non-PLS Islamic Finance**

**Murabaha:** Mark up sale, i.e., a sale on mutually agreed profit; client requests the bank to purchase an item for her; bank resells it to the client for a predetermined (‘marked up’) price usually paid in installments.



**Ijara/Ijara wal Iqtina:** Lease/lease purchase.

**Bai ‘Salam** (including **Istisna’a**): deferred delivery purchase; only items that can be fully specified in terms of quantity, quality, attributes, etc, are eligible; monetary instruments are specifically excluded.

**Bai’ Muajjal:** Deferred payment/credit sale, also known as *Murabaha Muajjal*, where seller informs buyer of cost, selling price and the final payment date (may be installment or lump sum); spot price may be lower than deferred payment price.

**Musawamah:** Normal market sale with no cost disclosure obligation.

**Jo’alah:** Service charge, consultancy fee, placement fee, etc.

**Qard-e-Hasanah:** zero-interest beneficence loans; repaid when/if able.

Source: El-Gamal (2000), Ayub (2002), Sundararajan and Errico (2002). See Ayub (2002) or Usmani (2002) for extensive discussion of these and other forms of Islamic finance.

### Selected References:

Abdel Haleem, M A S, translator. (2004) *The Qur'an*. Oxford, UK: Oxford University Press.

AMEinfo. (2006) *Corporate Sukuk Market Growing in the Middle East*. [www.ameinfo.com/96842.html](http://www.ameinfo.com/96842.html)

Ayub, Mohammed. (2002) *Islamic Banking and Finance: Theory and Practice*, Karachi, Pakistan: State Bank of Pakistan.

Bahrain Monetary Agency. (2006) “Banking Statistics”. [www.bma.gov.bh/cmsrule/media/pdf/statistics/bulletin/July2006small.pdf](http://www.bma.gov.bh/cmsrule/media/pdf/statistics/bulletin/July2006small.pdf)

Benaissa, Nasr-Eddine; Mayank P. Parekh; and Michael Wiegand. (2005) “A Growth Model for Islamic Banking,” *The McKinsey Quarterly*, October.

[www.mckinseyquarterly.com/article\\_page.aspx?ar=1694&L2=10&L3=51](http://www.mckinseyquarterly.com/article_page.aspx?ar=1694&L2=10&L3=51)

Chapra, M. Umer. (2000) “Is it necessary to have Islamic economics?” *Journal of Socio-Economics*, Number 29, pp. 21-37.

El-Gamal, Mahmoud Amin. (2000) *A Basic Guide to Contemporary Islamic Banking and Finance*. Plainfield: Islamic Society of North America. Available online at [www.ruf.rice.edu/~elgamal/files/primer.pdf](http://www.ruf.rice.edu/~elgamal/files/primer.pdf)

El-Gamal, Mahmoud Amin. (2003) “‘Interest’ and the Paradox of Contemporary Islamic Law and Finance.” *Fordham International Law Journal*, Volume 27, Number 1, pp. 108-149.

El-Hawary, Dahlia; Wafik Grais and Zamir Iqbal (2004) *Regulating Islamic Financial Institutions: The Nature of the Regulated*, World Bank Policy Research Working Paper #3227, Washington, DC: World Bank.

Euromonitor International. (2005) *Investment Banking and Venture Capital in the USA*, October. [www.majormarketprofiles.com/report\\_summary.asp?docid=42121](http://www.majormarketprofiles.com/report_summary.asp?docid=42121)

Friedman, Milton. (1969) *The Optimum Quantity of Money and Other Essays*. Chicago, US: Aldine Publishing Company.

Gassner, Michael Saleh. (2005) “Reasons to Issue Sukuk and the Structures Behind Them,” *Banker Middle East* Issue 56, (February) [www.bankerme.com/bme/2005/feb/how\\_to\\_structure\\_sukuk.asp](http://www.bankerme.com/bme/2005/feb/how_to_structure_sukuk.asp)

Glaeser, Edward L. and José Scheinkman. (1998) “Neither a Borrower nor a Lender Be: An Economic Analysis of Interest Restrictions and Usury Laws,” *Journal of Law and Economics*, Volume 41, Number (April) pp. 1-36.

Gulfnews.com. (2006) *Islamic Bonds Could Enter World Stage*. Reuters report.

- Humsawi, Roziana. (2005) "Demand for Sukuk Poised to Surge," *Business Recorder*, Malaysia (December 1<sup>st</sup>) [www.iifm.net/news3.php](http://www.iifm.net/news3.php)
- Healy, Patrick O. (2005) "For Muslims, Loans for the Conscience," *New York Times*, (August 7<sup>th</sup>).
- Islamic Development Bank. (2004) *29<sup>th</sup> IDB Annual Report 1424H (2003-2004)*. Jeddah, Saudi Arabia: Islamic Development Bank; calculated from Table 3.3 "Approved Project and Technical Assistance by Mode of Financing" [www.isdb.org/29annualreport/Chapter%2003.pdf](http://www.isdb.org/29annualreport/Chapter%2003.pdf)
- Khan, Feisal. (2006) *Pakistan: Islamic Banking by Judiciary?* Paper presented at the *Middle East and Central Asia Conference*, University of Utah, USA.
- Khan, L. Ali. (2003) "The Reopening of the Islamic Code: The Second Era of *Ijtihad*." *University of St. Thomas Law Journal*, Volume 1, Number 1 (Fall) pp. 341-385.
- Kuran, Timur. (2004a) *Islam & Mammon: The Economic Predicaments of Islamism*. Princeton, NJ: Princeton University Press.
- Kuran, Timur. (2004b) "Why the Middle East is Economically Underdeveloped: Historical Mechanisms of Institutional Stagnation." *Journal of Economic Perspectives*, Volume 18, Number 3, pp. 71-90.
- Mishkin, Fredric. (2007) *The Economics of Money, Banking and Financial Markets*. 8<sup>th</sup> edition. Boston, MA: Addison-Wesley.
- Parker, Mushtak. (2006) "Dow Jones Citigroup Sukuk Index Launched," *Arab News* (May 22<sup>nd</sup>). [www.arabnews.com/?page=6&section=0](http://www.arabnews.com/?page=6&section=0&article=82568&d=22&m=5&y=2006&pix=business.jpg&category=Business)
- Posner, Richard A. (1980) "A Theory of Primitive Society, with Special Reference to Law," *Journal of Law and Economics*, Volume 23, Number 1 (April) pp. 1-53.
- Presley, John R. and John G. Sessions. (1994) "Islamic Economics: The Emergence of a New Paradigm," *The Economic Journal*, Volume 104, Number 424 (May) pp. 584-596.
- Rahman, Fazlur. (1964) "Riba and Interest." *Islamic Studies* (Karachi) 3(1) Mar. 1964:1-43. English translation by Mazheruddin Siddiqi of the original "Tahaqiq-i-Riba," *Fikr-o-Nazar*, (Karachi) November 1963.
- Rammal, Hussain G. (2004) "Financing Through Musharaka: Principles And Application," *Business Quest: Journal of Applied Topics in Business and Economics*. [www.westga.edu/~bquest/2004/musharak.htm](http://www.westga.edu/~bquest/2004/musharak.htm)
- Sundarajan, V, and Luca Errico. (2002) *Islamic Financial Institutions and Products in the Global Financial System: Key Issues in Risk Management and Challenges Ahead*. International Monetary Fund Working Paper #192. Washington, DC: The International Monetary Fund.
- Udovitch, Abraham. (1975) "Reflections on the Institutions of Credits and Banking in the Medieval Islamic Near East," *Studia Islamica*, Numer 41, pp. 5-21.
- USC-MSA Compendium of Muslim Texts. Translation of Sahih Muslim, Book 10: *Kitab Al-Buyu'*. (Book of Transactions). [www.usc.edu/dept/MSA/fundamentals/hadithsunnah/muslim/010.smt.html#010.3613](http://www.usc.edu/dept/MSA/fundamentals/hadithsunnah/muslim/010.smt.html#010.3613)
- Venardos, Angelo M. (2005) *Islamic Banking and Finance in Southeast Asia:*



- Its Development and Future*. Singapore: World Scientific Publishing Company.
- Visser, Wayne A. M., and Alastair McIntosh. (1998) "A Short Review of the Historical Critique of Usury," *Accounting, Business & Financial History*, Volume 8, Number 2 (July) pp.175-189. Available online at [www.AlastairMcintosh.com](http://www.AlastairMcintosh.com).
- World Bank – International Monetary Fund. (2004) *Pakistan: Technical Note: Condition of the Banking System*. Washington, DC: World Bank & International Monetary Fund.
- Usmani, Mohammed Taqi. (2002) *An Introduction to Islamic Finance*. Arab and Islamic Law Series, The Hague, The Netherlands: Kluwer Law International.
- Warde, Ibrahim. (2000) *Islamic Finance in the Global Economy*. Edinburg, UK: Edinburgh University Press.
- Zaher, Tarek S, and M. Kabir Hassan. (2001) "A Comparative Literature Survey of Islamic Banking and Finance". *Financial Markets, Institutions and Instruments*, Volume 10, Number 4, pp.155-199.

*Feisal Khan*  
*Dept. of Economics*  
*Hobart and William Smith Colleges*  
*Geneva, New York*  
*khan@hws.edu*

## Labor Conditions, Hours and Rights

*John Burgess*

### **Introduction**

The employment relationship encompasses more than an exchange of labor services for wages. Behind the employment relationship are rights and obligations for both parties, often articulated through labor codes and employment laws, together with forms of non-wage benefits and issues of control and dependency associated with hours of work and working conditions. In turn the workplace is subject to regulations, the labor process is contested and contracts and collective agreements attempt to list conditions of employment and establish processes associated with dispute resolution.

Labor services cannot be separated from the worker who provides those services. In the extreme the individual can be commodified and subjugated by the buyer of the services—slavery involves the full possession of the services and the person providing the services. As such the purchase of labor services has to satisfy more than the conditions associated with a commodity exchange contract. The person providing the services will generally have rights of citizenship and political rights, they will also have needs as a person for rest breaks, meal breaks and time to fulfill other obligations such as caring for a family. Hence, work and the workplace is intensely political, behind the abstract market exchange can be found a complex set of rights and needs that are often subject to protective legislation, into workplace agreements or embodied in other forms of monitoring and regulation, including codes of conduct. In addition, the wage is one part of the reward system associated with employment. In many countries there are a range of statutory

employment entitlements including holidays and social security contributions. Finally, there are other qualitative aspects that affect the reward system associated with employment, these include the ability to control working hours and the pace of work, the intensity of work and the social relations of work. These aspects collectively contribute to job quality.

This essay will examine employment conditions, especially the hours of employment, and the broad set of conditions that impact on the quality of work. It will then examine labor rights and labor standards, and consider the other codes of conduct associated with developing civilised employment conditions across countries. Many of these issues have been central to the liberalisation of trade and to the development of trading blocks. At issue is whether labor rights and conditions represent an obstacle to free trade or whether they should be codified and enshrined in free trade agreements and form a non-negotiable foundation from which free trade can develop.

### **Labor Conditions**

Associated with employment is the expectation that workers will also have access to certain rights and conditions. Under the heading of rights are those civil and political conditions that are proclaimed by international organizations, including the United Nations and the International Labor Organisation (ILO). Such rights as the right to organise and the right to withdraw labor are covered by the covenants of labor standards that are administered by the ILO (discussed later). There are also other employment conditions that are enshrined through statutory employment norms, these include the right to holidays and other forms of leave and to social security coverage such

as pension contributions. These non-wage rewards vary across countries and access is often dependent on the nature of the employment contract and the legislative context in which work is carried out. In general, many non-wage employment conditions are confined to employees who are working full time and have on going employment arrangement. Part-time workers may have to establish a minimum threshold of hours in order to secure access while casual and temporary workers may not qualify for non-wage employment entitlements (Campbell & Burgess 2001). The self-employed do not have an employment contract or an employer, hence they are not subject to the rights and entitlements associated with employee status.

As forms of 'non-standard' employment arrangements proliferate across the OECD there is concern that more and more workers are falling outside of the legislative and collective net that offers forms of protection in employment and access to basic non-wage benefits. This is occurring as more workers fall outside of full-time, ongoing employee arrangements into a multitude of disparate employment arrangements including contract work, part-time work, temporary work, agency work, self-employment and traineeships (Burgess & Strachan 1999; Delsen 1999; Nolan 1999). What these forms of work do share to differing degrees is their exclusion from forms of protection and non-wage benefits.

In addition to the non-wage entitlements there are other non-wage aspects of employment that have an impact on working conditions. These include the ability to work in a safe and protected workplace, the ability to combine work and family responsibilities and the ability to work without being harassed, bullied or discriminated against. In

many countries there are extensive occupational health and safety codes that apply to workplaces and anti-discrimination legislation in place. Having the codes and the legislation does not necessarily mean that these conditions are enforced, as with all codes and regulations identifying their existence is different from having an enforceable regime of regulations and entitlements. – see the job quality section for a discussion of the full range of possibilities.

### **Working Hours**

The duration of working hours, the ability to control and vary working hours and the fluctuation in working hours can have an important impact on earnings and access to non-wage employment conditions including social security, career paths and training. Working time regimes have been restructured through workforce, technological, industrial and regulatory changes. On the supply side there are many job seekers who desire part-time hours for reasons of caring responsibilities, study or lifestyle choice. Equally, on the demand side employers are seeking a more precise temporal match between labor supply and labor demand in the face of fluctuations in labor demand over the day, the week, the month and the seasons. At one extreme there are those who work very short hours, at the other there are those working very long hours. Associated with each of these extremes are those who desire additional hours at the short hours end of the spectrum and those who desire fewer hours at the long hours end of the spectrum.

With the proliferation of both long and short hours jobs has come an expansion in contingent employment arrangements associated with irregular and fixed term work engagements. These range from contracting through to temporary

employment arrangements and multiple-job holding. Official labor force data tends to miss the diversity of employment arrangements and the proliferation of short-term employment arrangements. Multiple job holders only have their primary job counted, hence second and other jobs (usually part time) are not included in the workforce count. In Australia around six per cent of the workforce are multiple job holders. In the case of temporary jobs, especially through temping agencies, the very short incidence of jobs (hours or days) means that these jobs will be missed in the monthly labor force count. A third area of under-reporting is associated with clandestine employment arrangements, generally for avoiding tax, social security, employment and immigration regulations.

In its survey of working hours, the OECD (1998) highlighted the following trends emerging across the member countries:

A reversal was emerging in some countries of the long-term trend of declining average working hours – over the 1990s there was a slight increase in average annual working hours in Finland, Iceland, Sweden and the USA

A growing incidence of part-time employment, especially for women workers. The average male part-time employment share was 6.6 per cent, for women it was 26.4 per cent. In some countries the incidence of female part-time employment is over 40 per cent—the Netherlands, UK, Switzerland and Australia

A growing diversity of hours in terms of the number of hours worked and their spread across the day and the week. In many countries there has been an expansion in unsociable working hours across evenings, weekends and public holidays. In Australia only 48 per cent of the workforce have a

working schedule that operates between Monday and Friday (Watson et al 2003:86).

There is a growing share of employees who desire fewer hours of work. This is not surprising given the trend in some countries for very long (over 45 hours per week) working hours. Very long working hours are dominated by male workers and a relatively high incidence of very long hours can be found in Japan, the USA, Australia and the UK.

There is a strong push by governments, employers and employees for more flexible working hours arrangements. The factors driving this push include cost and productivity issues, lifestyle issues and work and family issues.

Hand in hand with workforce developments away from standard employment, there are emerging working-time arrangements that deviate from the “standard” working week as more and more workers are employed for either short or very long hours. In turn working hours encroach more into evenings and weekends as product markets are deregulated and production becomes continuous. This means more unsociable working hours arrangements and in some countries it has been associated with a rise in unpaid overtime hours (Campbell 2002). With working hour flexibility has come greater diversity and polarisation in working hours (Watts & Burgess 2001), and in many countries there is an active debate over how this flexibility can better fit other life commitments, especially family responsibilities (Bittman & Rice 2003).

In some countries there has been an active program to regulate working hours, notably France. This recognises the potential for job creation associated with capping working time and it also offers possibilities for considering other aspects of time use,

apart from work. The European Union has developed a working time directive as part of its social code (Leopold & Karsten, 1997) that sets out maximum weekly working hours (48), minimum rest periods per 24 hours (11 hours) and minimum periods of paid annual leave (4 weeks).

A central research question is the relationship between working time, productivity and employment. There remains the possibility of redistributing employment and reducing unemployment, yet retaining productivity levels, within a regime that places a cap on working time, such as in France. Bosch (2001) surveyed the effects on employment of collectively bargained working time reductions in selected EU countries. He found that there is no consistency in the results with the employment effects very dependent upon the type of plant (industry) and occupational category studied. In general there were positive employment effects, but this was within a context of collectively agreed working time reductions.

### **Job Quality**

In the main the issue of job quality is neglected in national and international discussions of the labor market, though through the research by the Economic Policy Institute in Washington (Mishel *et al* 2001) and the auspices of the EFILWC (2002) and the ILO (1999) it is becoming more prominent.

Europe has been active in developing directives for minimum rights and conditions at work, specifically in the areas of parental leave, and non-discrimination against part-time and temporary workers. The EU's employment strategy gives emphasis to the quality of work. Under the 2001 EU employment guidelines there is an emphasis on improving job quality in areas

such as working-time arrangements, job security and career breaks (EFILWC, 2002). While it appears that most of the debate and discussion is directed towards working-time arrangements the commitment to job quality should not prevent developments in other realms of work that impact on quality. Indeed, the EU has released a paper on the quality of work that outlines the challenges facing constituent governments in improving job quality across four key areas: career and employment security, health and well being, reconciliation of work and non-work activities, and skill development (EFILWC, 2002).

The ILO has a long agenda in addressing job quality through its program of minimum labor standards and rights at work. It has also highlighted what it calls as "the decent work deficit" (ILO 1999, 2001). The following components that constitute the decent work deficit are identified:

- *The employment gap*: those without jobs cannot actively participate in or progress within their communities;
- *The rights gap*: for those whose basic human rights are abrogated through work via such practices as bonded labor and child labor;
- *The social protection gap*: many jobs carry no protection from income loss through sickness, old age or childcare;
- *The social dialogue gap*: many workers have no voice in their dealings with governments and employers. They are often faced with impediments and legal constraints in collectively organising.

### **Indicators of Job Quality from Canada**

The Canadian Job Quality project is managed by the Work Network of Canadian Policy Research Networks. Sponsors of the web site are in the main public institutions (e.g. Alberta Human Resources and

Employment, and British Columbia Public Service Employee Relations Commission), but it also includes private corporations (e.g. Syncrude), charitable trusts (e.g. Atkinson Charitable Foundation) and trade unions (e.g. Canadian Labor Congress). The aim of the project is to go beyond the aggregate indicators of earnings and employment arrangements to look at the daily experience of work. The project provides a range of indicators on job quality and in turn disaggregates the evidence for each indicator by region, industry and workers characteristics. The purpose of the project is to “stimulate informed debate about the quality of people’s work, and to identify strategies on how to improve the quality of Canadian jobs” ([www.jobquality.ca](http://www.jobquality.ca)).

The indicators are grouped under 11 headings or themes (see Table 1). There is more than one indicator selected for each of these themes. Implicit in the headings is the belief that these are each important components of job quality. However, there appears to be no discussion of the reasons for the selection, neither of the criteria nor of the indicators. Also there is no rank ordering of each of the groups or themes, one is not regarded as being more important than the others are. There is no guide as to interpretation of the indicators; implicitly one gains a holistic view of job quality in Canada. In general, disaggregated data is provided to allow comparison across sectors and different groups of workers (e.g. by age and gender) re the distribution of job quality. The presented data represents the current state of job quality and there is no attempt to track the changes in or the nature of job quality through time. As expected, there is over-lap between the different themes; e.g. communications and unions at the workplace, and job demands and work schedules. There is no established minimum

criteria differentiating good from bad jobs, the framework is designed to allow for relative comparison and to identify gaps in jobs with respect to the different aspects of job quality. In a sense the project enables the benchmarking of jobs through their qualitative characteristics. It is also noticeable that collective coverage and voice is regarded as a special indicator as if there were a lack of agreement on whether this is a desirable job attribute. Finally, it is not a static framework; the homepage clearly asks for feedback and comments as well as suggesting that the framework is undergoing continuous upgrading and change in the light of user feedback.

Table 1 sets out all the job quality themes together with the indicators attached to each theme. What is striking is that the list of themes is extensive and that in total there are a large number of indicators used. However, while some themes have several indicators (e.g. communications and influence), others have few indicators (e.g. personally rewarding work). The information required to complete the job quality table ranges from the verifiable official data (e.g. temporary work, trade union density) to impressionistic and personal data associated with attitudes and perceptions. As such it is difficult to authenticate some of the indicators and to make cross job comparisons based on impressionistic views of work and the workplace.

### **Labor Rights and Labor Standards**

Labor standards are a collection of minima employment conditions associated with work. Individual countries may maintain standards that are well above the minimum. There is no rank ordering of the standards. The standards may be implicit not explicit, that is, they may not be legislatively enforced.

**Table 1 Themes, Indicates and Explanation of Canadian Job Quality**

<b>Job Quality Theme</b>	<b>Indicators</b>	<b>Explanation</b>
Communications and influence	Employee participation Employee influence as a workplace strategy Information sharing Receiving feedback on work	To have meaningful input into job design, work and the workplace.
Personally rewarding work	What workers want in a job	Factors that improve job satisfaction and non-pecuniary rewards from work.
Job security	Temporary employment Job security	Stability and predictability of employment in order to plan for holidays, seek loans etc
Job design	Job rotation Work teams	Work should be interesting, meaningful and allow for ability and skill development
Environment	Workplace health Workplace safety	OH&S in the workplace, good social relations at work and reasonable workloads.
Work schedules	Shift work Compressed work week Flexible hours Work hour preferences	Being able to control work schedules, work location and working time.
Relationships	Workers relations with supervisor Trust at the workplace	Getting along with managers/supervisors and fellow workers.
Job demands	Workload Long hours of work Unpaid overtime Work-life balance	Working a reasonable number of hours and having manageable workloads. Access to over-time pay and choice over working hours.
Pay and benefits	Profit sharing Merit based pay Satisfaction with pay	Extent of low pay, extent of reward satisfaction and diversity with respect to remuneration arrangements.
Training and skills	Skill use Opportunities for skill development Workplace training Access to training	The presence of opportunities to use skills and to develop new skills.
Other indicators	Union membership Unions, unpaid overtime, wages and benefits Other types of collective organisation	Unions within the workplace – extent of voice and representation; PC's in the workplace and the effect on working conditions.

*Source:* Adapted from [www.Jobquality.ca](http://www.Jobquality.ca)

Assessing whether standards are breached is difficult as is the enforcement of standards. Nevertheless our interest in labor standards has been heightened by the regular cataloguing of breaches of fundamental human rights often linked to conditions of employment (including forced labor/slavery, child labor, forced prostitution and hazardous employment conditions), not only in Africa or Asia (Evans 1995), but in the developed Western countries, including Britain and the USA (Brenchley 1996a,b; Lennon 1996).

The issue of labor standards in shaping and being an integral component of international trade policy was on the policy agenda of the League of Nations in 1927 and continues to be on the international policy agenda as the world moves towards extended

trading blocs and a more open international trade regime (Charnovitz 1987; OECD 1994). It has also been part of the charter of the International Labor Organisation since its origins in 1919 (Myrdal 1994). The protection of labor standards has in the past justified the application of a variety of trade barriers by developed countries mainly against those developing countries who were deemed to contravene acceptable labor standards (Herzenberg et al 1990:11; Singh 1990:252). Labor standards may involve many constituent clauses, but can be divided into national labor standards which establish minimum employment conditions within a country and international labor standards that establish a universal labor standard benchmark across all countries as part of a



trade regime that incorporates "acceptable" human rights and minimum labor standards. An international conference on labor standards (Herzenberg et al, 1990) adopted a working definition of labor standards that was country specific and incorporated minimum acceptable conditions, including the right of association, the right to organise and bargain collectively, the prohibition of forced labor, a minimum working age and acceptable working conditions.

Overlapping these standards are the conventions established by the International Labor Organisation (ILO) for signatory nations. There are over 170 such clauses or conventions. These include provisions governing fundamental human rights, minimum wages, employment security, working conditions and collective rights (OECD 1994:139). The enforcement of the standards is then the responsibility of those countries who ratify the standards. The ILO conventions do not set out punitive measures (eg trade restrictions) but are based on the goodwill of signatory nations. The effectiveness in implementing the standards in part depends upon obtaining ratification from nations for the standards. For example, in relation to 18 clauses setting out standards for wages and working conditions the extent of ratification ranged from 14 clauses for Spain down to two clauses for the United Kingdom and no clauses for the USA (OECD 1994:141).

The ILO has fundamental conventions that it regards as critical and are compatible with fundamental human rights. There are core conventions that include (Nyland 2000):

- Freedom of association and protection of the rights to organise (convention 87)
- The right to organise and to collectively bargain (convention 98)
- The suppression of forced labor (conventions 29, 105)

- Anti-discrimination in employment (convention 111)
- Equal remuneration for men and women for work of equal standing (convention 100)
- Abolition of child labor (convention 138)

In addition to the above trading blocs may also establish their own labor standards for member nations. A good example here is the European Union social charter which includes: procedural rules and consultation in the case of mass dismissal, the requirement that all contracts of employment be in writing, minimum occupational health and safety requirements at the workplace and a minimum 48 hour working week (OECD 1994:140).

There are regular reports of labor standards being breached across many countries. Where poverty is intense or the rule of law is limited in its application, labor and human rights are frequently breached. The ILO reports that up to 10 million children, the majority females, are working as domestic servants. They are frequently denied basic rights, receive no schooling and are subject to systematic abuse (ILO 2004). In some countries there have been attempts to legally end forced child labor in such industries as quarrying (Nigeria) and the circus (India) (Anti Slavery Society 2004a). The UN reported that regular abductions of women and children are taking place in Sudan, with many school children being forced to join militias (Anti Slavery Society 2004b). The International Confederation of Free Trade Unions (ICTFU) reported that over 200 trade union activists were murdered in 2002, with nearly all killings taking place in Latin America (ICTFU 2004). Columbia was cited as the most dangerous place on earth to be a trade unionist, with 184 trade unionists being murdered. The ICTU reported that export



processing zones, under the banner of free trade, had become bastions for anti union activity across many countries (ICTFU 2004). In some countries, while labor rights were proclaimed, they were often not enforced (Afrol News 2004). Forced and slave labor persists in many countries, with one report claiming that up to 25 thousand workers in Brazil are being pressed into clearing the jungle and producing pig iron from charcoal burners (Brown 2004).

There has always been a tension between uniform standards across countries as opposed to individual country standards (Okogwu 1994). Some commentators claim that Western labor standards disadvantage growth, development and living standards in poorer countries, where, for example, child labor, hazardous working conditions and long working hours are a matter of survival, not choice (Fields 1990). Other countries claim that labor standards represent “Western” values and are individual based, as such they ignore the collective, community and family/tribal affiliations that apply in other communities (Nyland 2000). In addition the enforcement of rights are difficult where, for example, child employment is the norm where families live in extreme poverty and there is an absence of available or adequate schooling (Pastore 2002).

### **Trade Liberalisation and Labor Standards**

Without international labor standards there will be pressure on the credibility and legitimacy of the trade liberalisation agenda associated with trade liberalisation. First, why should countries legitimise the fundamental abuse to human rights encountered through such work practices as forced labor, slavery and no collective rights? If expanded trade is to be a civilising force then such trade should conform to a set of core rules and

fundamental human rights. Second, without an agreed code of labor standards there is the risk of unilateral protectionism being generated by countries who wish to validly and invalidly exclude trade in commodities that are deemed to contravene acceptable international labor standards and human rights.

The issue of labor standards has come more to the fore in the context of the formation of trade blocs and the move towards a more liberal international trading regime (OECD 1994). Both sets of circumstances have led trade unionists and politicians to question whether labor standards can be sustained in such a context (Singh 1990). Globalisation confronts labor standards in three ways. First, the process of "labor flexibility" places pressure on existing labor conditions (Standing 1999). Under the banner of "international competitiveness" workers are being asked to compromise long-held employment conditions and rights. Examples include the extended spread of working hours together with the increase in unsociable working hours, the extension in fixed-term contract employment and the reduced collective representation of employees (OECD 1994). Second, the mobility of capital and its ability to move "offshore" further forces labor and individual governments to accept a choice between a dilution of employment conditions or a cessation of employment. Competitive bidding between countries for international investment often involves guarantees that erode employment rights and conditions (Marshall 1994). Third, globalisation has resulted in government policy shifting away from supporting full-employment towards satisfying the financial markets in the pursuit of inflation first and high interest rate strategies. This places downward pressure on wages and employment conditions and

maintains a large pool of surplus labor to maintain such pressure.

Establishing labor standards involves defining a core floor of rights and conditions that citizens could be expected to enjoy within a civilised community. Without these base rights there would be no limit to the depths that degradation and exploitation could sink to. The neo-classical view assumes that labor conditions will be elevated through trade liberalisation by market forces, intervention will only impede the levitation process. Those who support labor standards argue that social progress should be codified, that without standards countries will feel bound by the market to limit any improvement in working conditions and the absence of standards opens up the possibility for protectionist measures to sustain domestic working conditions (Emmerij 1994). The real problem is that trade liberalisation and market deregulation creates a downward pressure on employment conditions and wages which in many Western economies shows no signs of abatement (Hepple 1994). Wage levels and employment conditions will always be too generous and not competitive as compared to wage levels and employment conditions somewhere else in the world. A further round of wage restraint or a further loss of some employment condition will always be required.

### **Other forms of Protecting and Enforcing Labor Standards**

If governments are reluctant to act then consumers are not, consumer and human rights pressure has already forced many national and international businesses to proclaim and label that production has conformed to internationally recognised environmental and labor codes (Cowe 1996b; Nowicka 1996). In addition, under the banner of corporate social responsibility some global

companies are taking action to develop codes across supply networks. The Dutch based electronics company, Phillips, requires that its 50 thousand worldwide components suppliers comply with a number of codes that cover child labor, health and safety protection, anti-discrimination and the right to join a trade union (Anonymous 2004). Other companies are resorting to triple bottom line accounting (economic, social and environmental reporting) as well as social responsibility reporting (Tschopp 2003).

Labor standards have come to prominence as a by product of trade liberalisation and the formation of trading blocs. To some extent the labor standards of other countries forms a smokescreen to hide the dilution of labor standards in Western countries; the clear decline in labor standards in the USA is a good illustration of the lack of credibility behind recent moves to link labor standards with trade liberalisation. The growth of non-standard employment arrangements in many countries (OECD 1993) leads us to the question of whether labor standards have been placed in too limited an operational framework. There is the problem of dealing with non-employees, many of whom could be working long hours under dangerous conditions for below poverty line returns, and there is the problem with those who cannot access employment since they are unemployed. The important question to raise is whether we should go beyond waged employment and consider citizenship and economic participation (Hepple 1994). Labor standards in a sense establish a code for employment providing certain rights and conditions for employees, the problem is how far such rights should extend beyond employees and beyond employment.

Furthermore, trade liberalisation raises other types of standards including fundamental human rights, product codes and

environmental conditions. Some believe that trade liberalisation is potentially environmentally catastrophic since it forces countries to compete on the basis of the lowest environmental standards and that given the considerable global benefits associated with environmentally sustainable growth there should be a direct linkage between trade liberalisation and base environmental codes (Muckenberger 1994).

In general, Governments are reluctant to take unilateral action or to link trade to standards; the advantages of free trade to particular sectors of the community are a strong bargaining chip. There are exceptions, with the US for example linking an extension on the quota for Cambodian textile exports in 2004 to the meeting of minimum labor standards. Industries and producers can introduce their own codes of conduct in order to ally civil rights and consumer group activities regarding their production and employment arrangements (Jenkins 2002). Consumers can be mobilised into action against products that can be demonstrated to subvert base standards, whether labor, human rights or environmental. While companies have outsourced production to low cost countries (textiles and footwear is a notable example), often with a poor record for policing labor standards and human rights (Rosen 2002), consumer, non government development agencies and environmental groups have been prepared to step in to fill the regulatory void to monitor and publicise developments, and in some cases garner community action (Shaw and Hale 2002; Wilshaw 2002). In general, many international companies may participate in offshoring activities that subvert labor standards, but they are reluctant to be caught out and they are very sensitive to adverse publicity and consumer product resistance (Cowe 1996a). Such

pressures will continue with more offshoring arrangements proliferating and the spread of bilateral and multilateral free trade pacts.

### **Further Research Issues**

Many of the above issues require further research and evaluation. In terms of hours, the factors driving the diversity of employment arrangements, especially part-time employment growth, is deserving of examination. Also of interest is the impact of regulations on hours, for example, the capping of maximum weekly hours, on employment and productivity. Of interest is the spectre of rising average full-time weekly hours in some countries after almost a century of falling average full-time hours. Is this a temporary phenomena or does it represent a return to long working hours? Job quality remain relatively under-researched. Apart from definitional and conceptual issues, there is the public policy interest in improving job quality for the majority of workers. What mechanisms are best suited to achieving this task? The issue of labor rights and their implementation remains a major public policy question. Importantly the debate continues over the suitability of labor rights for poor countries and the link between the free trade agenda and labor rights and labor standards. The invigilation and implementation of labor standards remains an important public policy issue as consumer, environmental and other non-government organisations are increasingly filling the void vacated by many governments.

### **Internet Sites**

Canadian	Job	Quality	Project
	<a href="http://www.jobquality.ca">www.jobquality.ca</a>		
Clean	Clothes		Campaign.
	<a href="http://www.cleanclothes.org">www.cleanclothes.org</a>		

European Foundation for the Improvement of Living and Working Conditions. [www.eiro.eurofound](http://www.eiro.eurofound)

Labour standards and labour rights. International Labour Organization. [www.ilo.org](http://www.ilo.org)

Sweatshop watch [www.sweatshopwatch.org](http://www.sweatshopwatch.org)

## Selected References

Afrol News. (2004) *Malagasy Labour Rights Inoperative*. [www.afrol.com/articles/12853](http://www.afrol.com/articles/12853)

Anonymous. (2004) Phillips Electronics Enforces Standards on Suppliers. *Wall St Eastern Edition*. March 24, p. A8.

Anti Slavery International. (2004a) *Nigerian Police Rescue 200 Trafficked Children*. NPR.

Anti Slavery International. (2004b) *Forced Labour and Slavery in Sudan*. [www.antislavery.org](http://www.antislavery.org)

Bittman, M. and J. Rice. (2003) "The Spectre of Overwork: An Analysis of Trends Between 1974 and 1997 Using Australian Time Use Diaries", *Labour and Industry*, Volume 12, Number 3, pp. 5-26.

Bosch, G. (2001) "Working time: from Redistribution to Modernisation", in P. Auer (Editor), *Changing Labour Markets in Europe*. International Labour Office, Geneva, pp. 55-114.

Brenchley, F. (1996a) "Trade Agreement Prompts Human Rights Row", *Australian Financial Review*, August 8.

Brenchley, F. (1996b) "Myths About Child Labour Confronted", *Australian Financial Review*, December 13.

Brown, P. (2004) "Report Says 25000 Slave Labourers Are Working in Jungle of Brazil", *Irish Times*, July 19.

Burgess, J. and Strachan, G. (1999) "The Expansion of Non-Standard Employment in Australia and the Extension of Employer's Control", in A. Felstead and N. Jewson (Editors), *Global Trends in*

*Flexible Labour*. Macmillan Business, London, pp. 121-140.

Campbell, I. (2002) "Snatching at the Wind? Unpaid Overtime and Trade Unions in Australia", *International Journal of Employment Studies*, Volume 10, Number 2, pp. 83-108.

Campbell, I. and J. Burgess. (2001) "Casual Employment in Australia and Temporary Employment in Europe: Developing a Cross-National Comparison", *Work, Employment and Society*, Volume 15, Number 1, pp. 171-184.

Charnovitz, S. (1987) "The Influence of International Labour Standards on the World Trading Regime", *International Labour Labour*, Volume 126, Number 5, pp. 565-584.

Cowe, R. (1996a) "Moves to Purge Child Labour", *Guardian Weekly*, October 16.

Cowe, R. (1996b) "Retailers Back Code for Third World". *Guardian Weekly*, October 13.

Delson, L. (1999) "Changing Work Relations in the European Union", in I. Zeytinglu (Editor), *Changing Work Relationships in Industrialised Economies*. John Benjamins Publishing, Amsterdam, pp. 99-114.

Emmerij, L. (1994) "Contemporary Challenges for Labour Standards Resulting from Globalisation", in W. Sengenberger and D. Campbell (Editors), *International Labour Standards and Economic Interdependence*. International Institute for Labour Studies, Geneva, pp. 319-328.

European Foundation for the Improvement of Living and Working Conditions. (2002) *Quality of Work and Employment in Europe: Issues and Challenges*. Dublin, EFILWC.

European Industrial Relations Observatory. (2001) *Working Developments and the*

- Quality of Work.*  
[www.eurofound.europa.eu/eiro](http://www.eurofound.europa.eu/eiro)
- Evans, K. (1995) "Press Gangs Target Children", *Guardian Weekly*, November 5.
- Fields, G. (1990) "Labour Standards, Economic Development and International Trade", in S. Herzenberg and J. Perez-Lopez (Editors), *Labour Standards and Development in the Global Economy*. Washington DC: US Department of Labour, pp. 19-33.
- Hepple, B. (1994) "Equality: A Global Labour Standard", in W. Sengenberger and D. Campbell (Editors), *International Labour Standards and Economic Interdependence*. Geneva: International Institute for Labour Studies, pp. 123-132.
- Herzenberg, S., J. Perez-Lopez and S. Tucker. (1990) "Labour Standards and Development in the Global Economy", in S. Herzenberg and J. Perez-Lopez (Editors), *Labour Standards and Development in the Global Economy*. Washington DC: US Department of Labour, pp. 1-16.
- International Confederation of Free Trade Unions. (2004) *ICTU Survey 2003: 213 Trade Unionists Assassinated or Disappeared Worldwide*. [www.ictfu.org/](http://www.ictfu.org/)
- International Labour Office. (1999) *Decent Work*. Report from the Director. Geneva: 87<sup>th</sup> Session of the ILO.
- International Labour Office. (2001) *Reducing the Decent Work Deficit*. Report from the Director General. Geneva: 89<sup>th</sup> Session of the ILO.
- Jenkins, R. (2002) "The Political Economy of Codes of Conduct", in R. Jenkins, R. Pearson and G. Seyfang (Editors), *Corporate Responsibility and Labour Rights*. Earthscan, London, pp. 13-30.
- Job Quality Canada. (2002) [www.jobquality.ca](http://www.jobquality.ca)
- Lennon, P. (1996) "Women Caught in the Slave Trade", *Guardian Weekly*, April 14.
- Leopold, J. and Karsten, L. (1997) "The European Working Time Directive and the Control of Working Hours", in P. Flood, N. Heraty, M. Morley and S. McCurtin (Editors), *The European Union and the Employment Relationship*. Dublin: Oaktree Press, pp. 165-184.
- Marshall, R. (1994) "The Importance of International Labour Standards in a More Competitive Global Economy", in W. Sengenberger and D. Campbell (eds.) *International Labour Standards and Economic Interdependence*. Geneva: International Institute for Labour Studies, pp. 65-77.
- Milne, S. (1996) "Child Workers Number 250 Million", *Guardian Weekly*, November 17.
- Mischel, L, Bernstein, J. and Schmitt, J. (2001) *The State of Working America 2000/2001*. New York: Cornell University Press.
- Muckenberger, U. (1994) "The Other Side of the Coin: Globalisation, Risk and Social Justice", in W. Sengenberger and D. Campbell (Editors), *International Labour Standards and Economic Interdependence*. Geneva: International Institute for Labour Studies, pp. 133-141.
- Myrdal, H-G. (1994) *The ILO in the Cross-fire: Would it Survive the Social Clause?* Geneva: ILO.
- Nollen, S. (1999) "Flexible Work Arrangements: An Overview of Developments in the United States", in I. Zeytinglu (Editor), *Changing Work Relationships in Industrialised Economies*. Amsterdam: John Benjamins Publishing, pp. 21-40.
- Nowicka, H. (1996) "Oxfam Presses Retailers", *Guardian Weekly*, May 26.
- Nyland, C. (2000) "Asian Values, Human Rights and International Labour Standards", *International Journal of*

*Employment Studies*, Volume 6, Number 1, pp. 157-174.

OECD (1993) *Employment Outlook*. Paris: OECD.

OECD (1994) *Employment Outlook*. Paris: OECD.

OECD (1998) *Employment Outlook*. Paris: OECD.

Okogwu, G. (1994) "Labour Standards Across Countries With Different Levels of Development", in W. Sengenberger and D. Campbell (Editors), *International Labour Standards and Economic Interdependence*. Geneva: International Institute for Labour Studies, pp. 145-157.

Pastore, J. (2002) "Labour Standards and International Trade: the Case of Child Labour in Brazil", in T. Hamani (ed.), *Universal Wisdom Through Globalisation*. Tokyo: Japan Labour Institute, pp 96-106.

Rosen, E. (2002) *Making Sweatshops*. Berkeley: University of California Press.

Shaw, L. and A. Hale. (2002) "The Emperor's New Clothes: What Codes Mean for Workers in the Garment Industry", in R. Jenkins, R. Pearson and G. Seyfang (Editors), *Corporate Responsibility and Labour Rights*. London: Earthscan, pp. 101-112.

Standing, G. (1999) *Global Labour Flexibility*. Basingstoke: Macmillan.

Tschopp, D. (2003) "Its Time for Triple Bottom Line Reporting", *CPA Journal of New York*, Volume 73, Number 12, p.11.

Watson, I., J. Buchanan, I. Campbell and C. Briggs. (2003) *Fragmented Futures*. Sydney: Federation Press.

Watts, M. and J. Burgess. (2000) "The Polarisation of Earnings and Hours in Australia Under a Decentralised Industrial Relations System: the Lessons for Economic Policy", *International Journal of Employment Studies*, Volume 8, Number 1, pp. 27-58.

Wilshaw, R. (2002) "Code Monitoring in the Informal Fair Trade Sector: the Experience of Oxfam GB", in R. Jenkins, R. Pearson and G. Seyfang (Editors), *Corporate Responsibility and Labour Rights*. London: Earthscan, pp. 209-221.

John Burgess  
 Newcastle Graduate Business School  
 University of Newcastle  
 Australia  
 John.Burgess@newcastle.edu.au



## Labor Market Liberalization

*Mehmet Odekon*

### **Introduction**

Labor Market Liberalisation refers to the removal of all price and non-price barriers and rigidities in the labor market that prevent the wage rate from reflecting labor's true scarcity cost and that bar flexibility and mobility in the labor market. It is one of the more controversial aspects of economic liberalization programs. Labor market liberalisation has been strongly supported by the International Monetary Fund (IMF) and the World Bank and has been an integral part of their policy recommendations to developing countries (IMF 2001, 2000; World Bank, various issues). However it has been equally strongly criticized by its opponents for contributing to the rising income inequality and poverty (e.g. see Barnett and Cavanagh 1994 and Greider 1997).

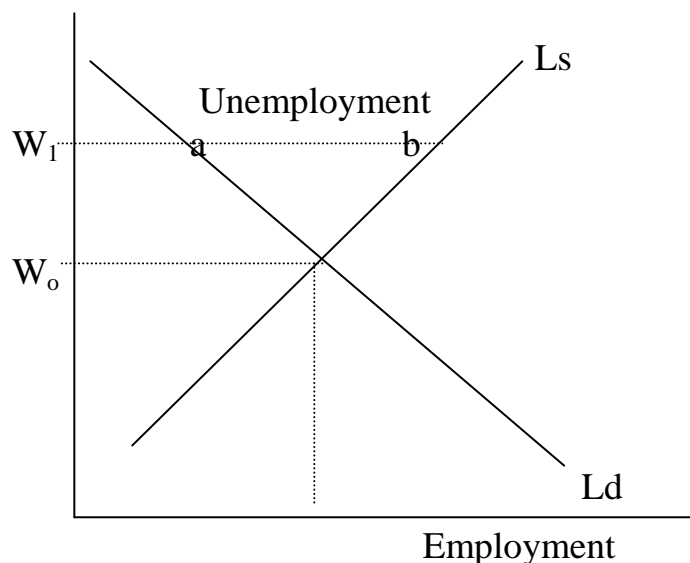
### **Economic Effects of Liberalisation**

A perfectly competitive labor market model assumes that there are a large number of firms in the market that are in competition with each other, and that the labor force is homogeneous- that is workers possess identical skills. In addition, it assumes perfect information and labor mobility, and that workers and firms do not exert control over the wage rate determined in the market. A perfectly competitive firm sets the real wage rate equal to the marginal product of labor (or it sets the nominal wage equal to the value of marginal product of labor) and finds the level of employment that would maximize profits. It should also be noted that in this framework, the diminishing marginal product of labor guarantees that the

wage rate be set to the lowest marginal product of the last worker hired, creating the opportunity for the producer to profit. Individuals' work-leisure decisions yield the labor supply.

The equilibrium wage rate ( $W_0$  in Figure 1) in the labor market in this model is one that clears the labor market. In other words, it sets labor demand equal to labor supply at that wage rate. In this competitive model, the equilibrium is unique and stable. The full flexibility of wages ensures that deviations from equilibrium are only temporary and that the market returns automatically to equilibrium as wages move up or down to eliminate the disequilibria. Thus, involuntary unemployment does not persist in this model. Voluntary unemployment, however, is the result of individual choice, preferring leisure to work.

*Figure 1: High Minimum Wage Rate*



If the reservation wage, the lowest wage at which someone would decide to work, is higher than the market determined wage, that person chooses leisure over work. In a technical sense there is no voluntary unemployment in the market. Every

individual who is willing to work at the ongoing market wage finds work and is employed. The model, as presented, makes sure that labor is efficiently allocated among alternative uses, implying that the total output in the economy is maximized, given the available labor supply.

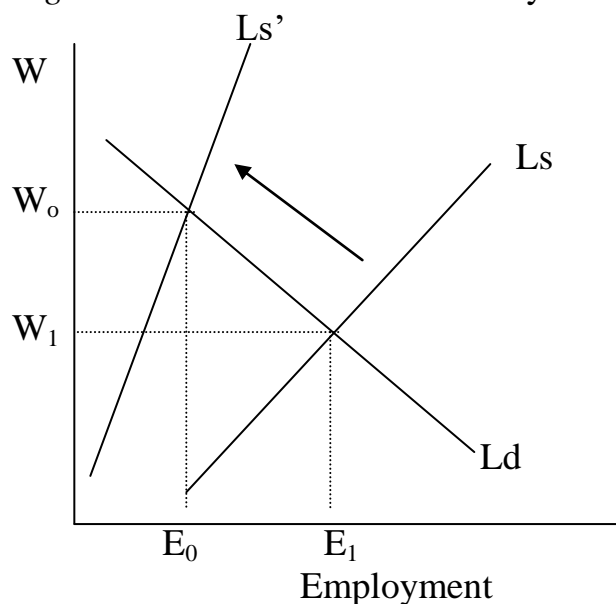
There are various factors that can contribute to a labor market's failure to function competitively. For instance, the presence of monopolies or monopsonies causes the wage rate to remain below the value of the marginal product of labor and the level of employment to decrease below the perfectly competitive level. In other cases, for example in countries with a minimum wage that is too high and/or with powerful labor organizations, the wage rate can be fixed at a level higher than the equilibrium wage rate, raising the cost of labor and also resulting in unemployment. In Figure 1, the equilibrium wage rate that clears the labor market is at  $W_0$ , corresponding to the point where labor demand and supply intersect. If the minimum wage rate is fixed at  $W_1$ , labor supply ( $L_s$ ) exceeds labor demand ( $L_d$ ), causing unemployment in the amount  $L_s - L_d$ .

As long as the wage rate is fixed at  $W_1$ , the system would discourage the profit-maximizing producer from hiring labor and encourage the producer to substitute capital for labor, perpetuating unemployment. Deregulation in the labor market would cause the wage rate to decrease to its equilibrium level  $W_0$ , as the unemployed bid the wages down. At  $W_0$ , unemployment is eliminated as work is provided for everyone who is willing to work at the wage rate  $W_0$ . Full wage flexibility in a perfectly competitive labor market, theoretically, would eliminate all disequilibria automatically and assure full employment at the equilibrium wage rate that would be

equal to the marginal productivity of the workers.

Similarly, a low level of labor mobility raises the cost of labor and leads to inefficient allocation of resources. Labor market immobility could result from discrimination, restrictive employment legislation, and also from geographic and other institutional factors in the market. In Figure 2, the lack of labor mobility causes the labor supply schedule ( $L_s$ ) to be relatively steep, resulting in a relatively high wage rate,  $W_0$ . Introducing measures to raise labor mobility would not only shift the labor supply schedule down to  $L_s'$  but would also make it relatively flatter because of efficiency gains. At the new market equilibrium, the wage rate is lower and the employment level higher.

*Figure 2: Labor Market Immobility*



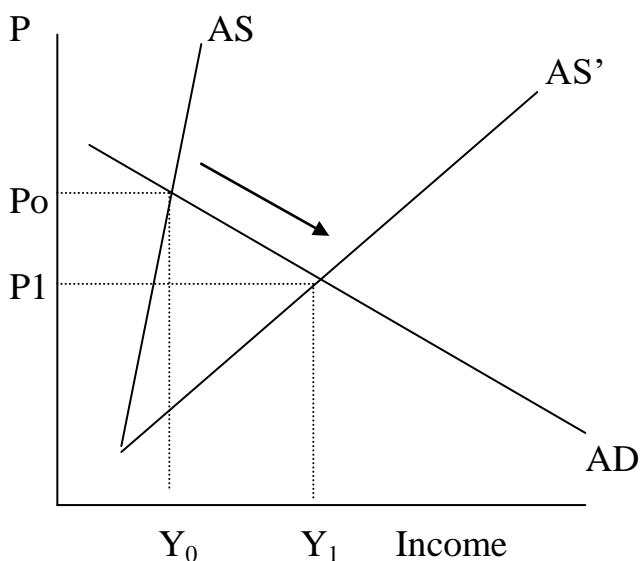
Thus, labor market liberalization along with the liberalization of output, other input, and external markets would lead to more efficient allocation of resources. In Figure 3, the economic effects of overall economic liberalization are captured. In the figure, economy-wide aggregate demand and supply determine aggregate price and



production in the economy. The pre-liberalisation aggregate supply schedule is given by the relatively steep  $AS_0$ , and the economy-wide aggregate demand is  $AD_0$ . The initial equilibrium values of aggregate price and output are  $P_0$  and  $Y_0$ , respectively.

The shift in  $AS_0$  to  $AS_1$  is partly the result of the reduction in the cost of labor and partly the result of the improvement in allocative efficiency in the economy. As a result, output and employment rise with stable prices. In this sense the economy overall is better off with economic liberalisation.

*Figure 3. Overall Economic Liberalisation*



### Neo-Liberalism and Labor Markets

Supporters of neoliberal supply-side economic policies in the 1980s, notably President Ronald Reagan and Prime Minister Margaret Thatcher, pressed successfully to incorporate labor market liberalisation and deregulation into the Washington Consensus, and hence into IMF and World Bank policy recommendations. In the 1990s, President Bill Clinton, Prime Minister Tony Blair, and Chancellor Gerhard Schroeder, inspired by Anthony Giddens' Third Way (1998 and 2000), deepened further labor

market liberalisation. Giddens argued that labor markets, like any other market, should respond structurally to increasing competition in the global marketplace. As Giddens (1998:65-66) puts it, one of the main pillars of the Third Way is the (ethical) principle, 'no rights without responsibilities.' It is this principle that later has led to the 'workfare, not welfare' dictum, trimming unemployment benefits as well as other pecuniary and non-pecuniary benefits, arguing that (1998:115) "benefits have their own autonomy, regardless whether or not they meet the purposes for which they were originally designed.

As this happens, expectations become 'locked in' and interest groups entrenched." The solution to this dilemma is to make labor contribute to their benefits, which at the same time reduces the non-wage cost of labor to the producer. These decreasing labor costs coupled with the dynamic market economy and 'entrepreneurial energy' would render firms more profitable and hence more willing to undertake new investment and to create new employment and income, reducing poverty and income inequality. The critics of the Third Way have attacked it for its blatant support for the market economy and for dismantling the traditional social safeguards without proposing new schemes (Barkan 2000).

### Globalisation Issues

Despite these expectations, the positive effects of labor market liberalisation have been limited. Labor market and overall economic liberalisations in a global economy are generally expected to lead to factor price convergence, reducing income inequality as it did during the late nineteenth century globalisation episode (Williamson 1998). Indeed, opening up to trade is hypothesized to reallocate resources into

production areas where a country has a comparative advantage. Over time, free trade and economic integration are said to equalize wages and other factor prices between trading partners, reducing intra- and inter-country inequality. However, in spite of significant labor market and trade liberalisations (and economic integration) during the current period of globalisation, convergence is not yet fully supported by empirical evidence (Feenstra 1998). In addition unemployment has increased, especially among unskilled labor.

The rise in income inequality and unemployment can be explained by the two unique characteristics of contemporary globalisation that are interrelated: technological innovation, especially the revolution in communication technology; and slicing the value chain (Krugman 1996). Both of these have introduced a so-called skill bias in the labor market. In the case of technological innovation, demand for skilled labor increased, raising its wage relative to unskilled labor and causing income inequality to increase. Slicing the value chain refers to the outsourcing of the production process, predominantly internationally. Outsourcing parts of the production process to various low wage countries has reduced demand for unskilled labor in industrial countries, further contributing to rising income inequality and unemployment. Feenstra and Hanson (1996) have shown that demand for unskilled labor is actually decreasing in both industrial and developing countries. Even though the outsourced process uses unskilled labor in the industrial country, it is the relatively cheaper skilled labor that is employed in the developing country. Thus, wage and income inequality in both countries rises, as does the skilled labor intensity in production. Outsourcing has been especially prominent

in textile, clothing, leather, and footwear industries. The receiving countries have traditionally been developing economies in Asia and East Asia. The search for the lowest possible labor cost countries for outsourcing is facilitated by means of electronic communication and falling transportation costs. This has enabled even the most distant developing economies to be part of the global outsourcing network.

Low pay and sub-optimal working conditions, disregard of occupational safety measures, and the use of child labor in developing countries that produce for multinational corporations have been subject of intense criticism, as have the loss of unskilled jobs in industrial countries (see Klein 1999; D'Mello 2003; AFL-CIO). Labor market liberalization and flexibility have increased the vulnerability and insecurity of unskilled labor, and have reduced the bargaining power of labor as a whole. The loss of union power has also contributed to income inequality (Borjas and Ramey, 1995) and has caused the burden of taxation to fall disproportionately more on labor, reducing the paycheck of workers (Rodrik 1997). Joseph Stiglitz (2002:84), Nobel laureate and former chief economist to the World Bank, refers to labor market flexibility as "simply a code name for lower wages, and less job protection." In order to assess the full impact of labor market liberalization one has to look at it within the context of overall economic liberalization. On one hand, unemployment among the unskilled is increasing and take-home wages of the employed are decreasing and job security is being taken away; on the other hand, the social support net is being weakened by welfare reforms. This erosion of the economic status of labor has also led to the erosion of the lower middle-income classes, almost universally (Vacs & Renwick

1998; Nas & Odekon 1998; Koechlin 1998). As globalization hits more and more groups hard, a globalization backlash - such as during the 1914-1945 period - could be possible. According to Williamson (1998:69-70): "a move towards de-globalization is unlikely to repeat. Yet history does supply that warning: if a globalization backlash can be found in our past, it may reappear in our future".

### **Liberalisation of Capital and Labor Markets**

The neoliberal school believes that full mobility of international capital is the engine for growth. The lender and the borrower are better off as a result of capital market liberalization and mobility.

Developing countries, in general, are characterized by not having enough domestic savings to finance the necessary investment needed for development. A low level of income leads inevitably to a low level of domestic saving, both private and public. One way of closing this investment-saving gap is by transferring financial foreign resources from abroad by borrowing or by attracting foreign direct and/or portfolio investment.

Prior to the 1980s, world financial markets were glutted with petro-dollars, and both bank and non-bank international financial institutions loaned large amounts to developing countries to reduce their excess foreign reserve (especially dollar) holdings. This exposed the financial institutions to high risk. By 1983, a critical number of developing countries found themselves in a position of not being able to service their debt. The resulting world debt crisis called for an examination of the efficiency and effectiveness of the capital markets by industrial countries. The reforms that followed emphasized the need for

liberalization in developing countries in such a way that they would attract foreign investment and would compete with other borrowers in international financial markets. The International Monetary Fund, the World Bank and the World Trade Organization became the overseers of the correct implementation of these reforms. In this context, liberalization meant restructuring the developing economies after the industrial economies by removing all the barriers to the free functioning of all markets, domestic and external, in the economy. The labor markets were not an exception to this rule. The liberalization of the labor market ensures that the price of labor reflects the true opportunity or scarcity cost of labor in the economy.

In this environment the lender and/or investor maximizes his/her return in a risk free manner, whereby the borrower obtains savings that is lacking domestically from abroad for investment. The borrower also enjoys gains in economic efficiency, as international capital brings new know-how, knowledge, and ideas. However, not everyone shares this win-win scenario.

The neo-Marxist school regards global labor market liberalization as a means of exploiting labor by capital (see Gindin 2002; Grabel 2002; Sweezy et al 2002). Within this framework, the cohabitation of rich and poor countries in the world economic system inevitably leads to unequal power relations between the industrial core and developing periphery countries. Outcomes from this relationship always favor the rich industrial countries. Accordingly, international capital moves across borders only in search of cheap labor to maximize profits and to avoid economic stagnation at home. As capital is invested and reinvested at home, profits start being squeezed, leading to the Marxist dilemma

between profits and wages. Thus, liberalisation of labor markets abroad makes sure that a wide selection of countries are continuously available for the international capital to flow in, in order “to take advantage of a compliant labor force and a less regulated business environment” (Grabel 2002:39). Neoliberal financial reforms and labor market liberalisation only aggravate inequality, favoring domestic and international capital over labor. As profits in one location decline, capital moves out, leaving behind an economy more dependent on global capital and more vulnerable to crises than ever before. Progressive non-Marxist economists, for instance Rodrik (1997), have criticized this speculative nature of global capital and recommend capital controls and/or a Tobin tax, which is levied to all foreign exchange transactions in the international financial markets. The driving force behind this and similar recommendations is that nations should have the ability to control capital flows thereby minimizing their adverse effects. A full list of proposed controls on finance capital and their discussion is available in ATTAC (2003).

### **Policy Recommendations**

The reallocation of resources in a global economy along the lines of comparative advantage needs flexible and competitive markets. Labor market liberalization indeed facilitates the reallocation of labor. However, if export industries are slow in absorbing the labor released from import-competing industries, the economy faces in the interim an adjustment cost in the form of unemployment. In a perfectly competitive environment the unemployment would only be temporary, and full employment would eventually be restored. In developing countries, specifically in those at the early

stages of development, the adjustment costs can mount very rapidly leading to mass discontent. In the industrial world, several countries have introduced measures to protect labor in import competing industries from losses resulting from international trade. In the United States, for instance, trade adjustment assistance provides income support to unemployed workers who lost their jobs because of cheaper imports. In addition, the U.S. trade law even allows temporary protection to import competing industries in order to enable them to become competitive. Similar measures in developing countries could be implemented to minimize the short-run effects of liberalization on labor markets.

One of the major effects of labor market liberalization has been the weakening of the position of labor relative to capital. The high mobility of capital has enabled it to escape national scrutiny, allowing income to be redistributed away from labor towards capital. As we have seen, liberalization hurt most low pay, unskilled labor. Since the wage gap is a direct outcome of the skill differentials, the human capital content of unskilled labor should be the primary public policy target. Provision of formal and/or informal education to enhance the skill and productivity of unskilled labor is essential. An additional way of reducing this income gap could be a tax and subsidy scheme, taxing international capital through a Tobin-type tax and providing a wage subsidy to low-pay unskilled labor. In developing countries, temporary measures similar to those in the U.S. could also be implemented to minimize the burden of the short-run adjustment costs to labor. Above all, however, labor organizations in developing as well as industrial countries need to become stronger and more politically active to obtain a strong bargaining position vis-à-

vis capital. The global capital lobby stretches from the WTO, the IMF and the World Bank to the World Economic Forum, and G7/G8 summits. These affect major economic, political, and social decisions and outcomes. Labor needs to organize globally and to assert itself as an equal partner to capital. As Faux argues, the existing set of rules in the global marketplace focus on protecting the investor and securing the return to capital. Therefore, “the main strategic task before us is the strengthening of the alliance of the working people ... through a common program” (2003:90). This common program should revolve around the central role labor plays in the production process. In this regard, trade unions provide labor the opportunity to collectively represent itself within the institutional framework of the world economy. According to a report presented by the Congress of South African Trade Unions (COSATU 2003) at the 2003 World Social Forum (an alternative to the World Economic Forum) the global solidarity of the labor force starts in the nation state with a genuine commitment to the principles of social justice and participatory democracy. This awareness should ultimately lead to the rebuilding of a democratic global trade union movement that considers itself a political and economic partner in the process of negotiating and shaping the evolution of the global society as well as the economy. Such a global coordination and networking by the trade unions can prevent the erosion of the economic status of labor vis-à-vis capital and strengthen the partnership between them.

### **Selected References**

AFL-CIO. <http://www.aflcio.org/issues/politics/globaleconomy/>

- ATTAC. (2003) “Financial Capital”, in Fisher, William and Thomas Ponniah (Editors), *Another World Is Possible: Popular Alternatives to Globalization at the World Social Forum*. London: Zed Books, pp. 41-50.
- Barkan, Joanna. (2000) “The Third Way/Die Neue Mitte”, *Dissent*, Spring, pp. pp. 51-65.
- Barnet, Richard and John Cavanagh. (1994) *Global Dreams: Imperial Corporations and the New World Order*. New York: Simon and Schuster.
- Borjas, George and Valerie Ramey. (1995) “Foreign Competition, Market power, and Wage Inequality”, *Quarterly Journal of Economics*, Volume 110, Number 4, pp. 1075-1110.
- COSATU, “Labour: A Strategic Perspective on the International Trade Union Movement for the 21<sup>st</sup> Century”, in Fisher, William and Thomas Ponniah (Editors), *Another World Is Possible: Popular Alternatives to Globalization at the World Social Forum*. London: Zed Books, pp. 62-74.
- D’Melo, Bernard. (2003) “Reebok and the Global Footwear Sweatshop”, *Monthly Review*, Volume 54, Number 9, pp. 26-40.
- Faux, Jeff. (2003) “A Global Strategy for Labor”, in Fisher, William and Thomas Ponniah (Editors), *Another World Is Possible: Popular Alternatives to Globalization at the World Social Forum*. London: Zed Books, pp. 78-90.
- Feenstra, Robert. (1998) “Integration of Trade and Disintegration of Production in Global Economy”, *Journal of Economic Perspectives*, Volume 12, Number 4, pp. 31-50.
- Feenstra, Robert and Gordon Hanson. (1996) “Productivity Measurement and the Impact of Trade and Technology on

- Wages: Estimates for the US, 1972-1990", *NBER Working Paper*, No. 6052.
- Giddens, Anthony. (2000) *The Third Way and its Critics*. Cambridge, UK: Polity Press.
- Giddens, Anthony. (1998) *Third Way: The Renewal of Social Democracy*. Cambridge, UK: Polity Press.
- Gindin, Sam. (2002) "Social Justice and Globalization", *Monthly Review*, Volume 54, Number 2, pp. 1-11.
- Grabel, Ilene. (2002) "Neoliberal finance in the Developing World", *Monthly Review*, Volume 53, Number 11, pp. 34-46.
- Greider, William. (1997) *One World, Ready or Not: The Manic Logic of Global Capitalism*. New York: Simon and Schuster.
- IMF (2001) *Social Issues In IMF Supported Programs*. Washington DC: IMF.
- IMF (2000) *World Economic Outlook: Supporting Studies*, IMF, Washington, DC.
- Klein, Naomi. (1999) *No Logo*. New York: Picador.
- Koechlin, Timothy. (1998) "Liberalization and Mexican Labor Markets", in Dabir-Alai, Parviz and Mehmet Odekon (Editors), *Economic Liberalization and Labor Markets*. Westport, CT: Greenwood, pp. 179-198.
- Krugman, Paul. (1996) "Does Third World Growth Hurt First World Prosperity?" *Harvard Business Review*, Volume 72, pp. 113-121.
- Nas, Tevfik and Mehmet Odekon. (1998) "Economic Liberalization and the Turkish Labor Market", in Dabir-Alai, Parviz and Mehmet Odekon (Editors), *Economic Liberalization and Labor Markets*. Westport, CT: Greenwood, pp 239-250.
- Rodrik, Dani. (1997) *Has Globalization Gone Too Far?* Washington DC: Institute for International Economics.
- Stiglitz, Joseph. (2002) *Globalization and its Discontents*. New York: Norton.
- Sweezy, Paul, Harry Magdoff, John Bellamy Foster, and Robert McChesney. (2002) "The New Face of Capitalism: Slow Growth, Excess Capital, and a Mounting Debt", *Monthly Review*, Volume 53, Number 11, pp. 1-14.
- Vacs, Aldo and Trudi Renwick. (1998) "Argentina: Neoliberal Restructuring and Its impact on Employment, Income, and Labor Organization", in Dabir-Alai, Parviz and Mehmet Odekon (Editors), *Economic Liberalization and Labor Markets*. Westport, CT: Greenwood, pp. 155-178.
- Williamson, Jeffrey. (1998) "Globalization, Labor Markets and Policy Backlash in the Past", *Journal of Economic Perspectives*, Volume 12, Number 4, pp. 51-72.
- World Bank (various issues) *World Development Report*. New York: Oxford University Press.

Mehmet Odekon  
Department of Economics  
Skidmore College, Saratoga Springs,  
New York State, USA  
modekon@skidmore.edu

## Labour Market Training Programs

*Eduardo Tomé*

### **Introduction**

This entry analyses the theories and practices concerning labour market training programs (LMTP) throughout the world. It is divided in four main parts. In the first part, the concept of LMTP is defined and discussed; furthermore, the main theories regarding LMTP will be reviewed and analysed; the analysis will be made focusing in the point of view of the main actors, that is: the trainees; the companies and organizations; the government and public authorities; the other types of entities: unions, trainers, families; and finally, the society as a whole.

In the second part, the main known experiences on LMTP will be described. The description will follow the following guidelines: type of LMTP; type of individuals or organizations supported; type of funding and administration; dimension (absolute and relative) of the program; analysis of the determinants of participation; socio-Economic evaluation. This part will end with a topic on the global or multinational significance of LMTP, which, given that we live in an increasingly competitive world it is quite interesting to address.

In the third part a critical analysis of the theories presented in the first part will be made, according to the data presented in the second part. This part will address three important questions. Are LMTPs worth? Should they continue to be made? Are there better ways than others of making LMTPs? Finally, we will consider the political and administrative problems regarding the evaluation of LMTPs.

In the fourth part the conclusions will be summarized; also suggestions for further research, and suggestions for potential governance innovations will be presented.

Nowadays, it is well understood that the world lives in a “Knowledge Era”. And, it is also considered that in this economic epoch Knowledge is the most important economic asset. Having said that, four very important notes must be made, on the relation between the Knowledge Era and the LMTPs:

First of all, historically, the Knowledge Era is in itself the corollary of a very complex evolution, in terms of use of the production factors in the production process. Specifically, until the 19<sup>th</sup> century, when Karl Marx wrote his well known books on the class struggle (*Capital*, and *Communist Manifesto*) the labour force was mainly very unskilled and the workforce of the farming sector and of the industrial sector was essentially defined by the number of arms available. Exception existed, like the Egyptians scribes or the navy captains of the Renaissance, but essentially, in that context, the opposition between the brute force of labour, and the power of the machines owned by the capitalists, made perfect sense.

In the beginning of the 20<sup>th</sup> century, after Frederick Taylor had made his analysis on the Scientific Organization of Firm, the labour force began to be more skilled (Taylor, 1964). In fact, Taylor defined levels of skills needed for each member of the labour force, according to the position he occupied in the organization; that was the idea that generated competence profiles. It is well known, that the Tayloristic organizations had a pyramidal structure of power, and therefore it was understood that most of the workers would need only very limited skills. But Taylorism gave LMTPs, education and training (E&T), and on-the-job training, a central role in the



organizations, because they decisively needed the competences and skills learned in those circumstances.

Afterwards, the 20<sup>th</sup> century saw, at least in the developed world an immense surge in competences of people and organizations. And in the fifties and sixties, Human Capital began to be considered as a fundamental economic asset, which was particularly decisive in order to take the most profit from the other production factors: energy, machines, installations, etc. That evolution was contemporary to the elaboration of a series of works that became known as the human capital theory (Becker 1993). Accordingly, in the eighties, in Central and Northern Western Europe, Japan, the USA, Canada, Australia and in Eastern Europe, the average years of schooling neared 10, and training was a well established reality.

In the Third Industrial Revolution, post-Taylor organizations emerged, where the pyramid of power was substantially inversed, and the number of competences needed by the average worker was greater than in Taylorist organizations. Therefore, from the beginning to the end of the 20<sup>th</sup> century LMTPs became more important. Finally, in the last decade of last century, with the advent of the internet, the world entered the Information Age in which knowledge is decisive. In this context LMTP gained a new meaning, (see next note) but continued to be economically fundamental.

Secondly, it is important to note that if LMTPs are not the only way of acquiring knowledge, they are an important one. The notion of knowledge is in itself a very broad one. For knowledge managers, information is organized data, and knowledge is “understood information” (Maurer, 1998). Knowledge workers are those that use knowledge intensively, and they use acquire knowledge from different sources. The most

important of those sources are the following: formal education and training, professional experience, on-the-job-training, informal learning, self training, research and development, scientific work, organizational routines. LMTPs may comprise some of those activities.

In third place, the importance of LMTPs may differ considerably in developed countries or in the less developed countries. In one hand, it is easy to consider that LMTPs are fundamental to give basic competences to people with low educational levels, and skills, as those that constitute the labour force in developing countries and poor countries. But, on the contrary, and in the other hand, in the developed world, given the average high levels of skills of the population, the formal LMTPs may be substituted or at least complemented, by more informal, short embedded and work related forms of acquiring knowledge.

Finally, in the world of the 21<sup>st</sup> century, it is well known that knowledge itself changes quite rapidly. Due to the need of specificity and efficiency, competence profiles become more complex. Nowadays, (April 2007) almost 1 200 000 entries are available on the subject through Google. Competence profiles are needed more than ever; but they are also being reviewed every year. Both the educational and training system that train people, and the organizations that hire them, have big doubts on what to teach to whom.

Even given what was previously said there is a strong case that LMTPs are still important to societies and relevant to the world around us.

## **Definitions**

Some important definitions may be made in the literature on the subject of this paper: LMTPs are organized structures of operations that provide support to



individuals and organizations in order to increase their chances in the labour market. LMTPs are usually divided in several parts: classroom training, on job training, job search, support to migrations, and direct job support. Crucially, from an economic point of view, LMTPs are investments. This means that LMTPs use scarce resources in order to fulfil needs.

The needs that are addressed through LMTPs relate to jobs. In some cases, it is important to improve the competences of workers and organizations; in other cases it is important to find a job to an individual, or to find the individual for a given job; finally, it may also be important to subsidize a job for a while. The investment made using LMTPs should increase the stock of Human Capital (Becker, 1963), Intellectual Capital (Edvinsson and Malone, 1997; Kaplan and Norton, 1994; Sveiby, 1998) Organizational Capital (Nonaka, 1994; Black and Lynch 2005) or by and large, of Knowledge (Maurer 1998), of individuals and societies. However; in some particular cases, like in the case if the support to job search or migrations, that investment should enable the individual to find a way of using the stock of Human, Intellectual or Organizational Capital that already exists.

LMTPs may be produced by private, public or third sector organizations. The public sector may have different types of roles regarding LMTPs: produce, fund, or legislate. LMTPs are evaluated by the evolution of outcomes that they imply to the individuals and the organizations involved: this is called the “partial equilibrium” approach; it is also possible to try to find the impact of LMTPs programs in the society: this is called the “general equilibrium approach”; when LMTPs are big enough in relative terms, they may have societal impacts. The impact of LMTPs is seen in

variables such as wages, employment, productivity, product quality, and exports.

In this context, a considerable difference may exist between the outcome as perceived by the person or organization that invests (called the private return) and as perceived by the society (called the social return). Given that knowledge may be easily transmitted, a positive difference exists frequently between the social return and the private return to LMTPs. This positive difference is called “positive externality”, and is a decisive rationale to base the public intervention in LMTPs.

## **Review of Literature**

In this section, the main economic and sociologic ideas about LMTPs are described. For the sake of clarity, the exposition will be divided in six parts. The first four parts related to each one of the relevant actors in the LMTPs: trainees and other supported individuals companies and organizations; government and public authorities; other types of entities, such as unions, trainers, and families. In the analysis of each actor, two main questions are discussed: participation rate and outcomes; both questions are analysed in a Cost Benefit Analysis framework. The fifth part addresses the relation between LMTPs and the society as a whole. A final subsection deals with the question of the evaluation of LMTPs fundamentally addressing the same two questions, namely, participation rates and outcomes.

### *The Case of Trainees and other Supported Individuals*

According to the Human Capital Theory individuals invest in education and training because they think they will benefit in the middle to long run. The cost associated with E&T is measured in time, forgone earnings,

leisure (particularly if the investment is made after working hours), tuition fees, and pedagogic materials. The benefit derived from E&T is measured in wages (Mincer 1962; Becker 1993), employability (Card & Sullivan 1988), or other variables like quality of work.

In the case of LMTPs, it is very important to make three points. First, the participation costs may be significantly reduced if the forgone earnings become negligible; this situation may happen in two main cases. These include if the participation in the LMTPs is made during the working hours, as a part of the working relation, without wages reduction or obligation to remain in the company after the training experience. It also includes if the participation in the LMTP is an after work activity and in this case the cost should be measured in terms of leisure activities and related aspects. Second, in the case of public sponsored LMTPs a grant may be awarded to the participants, especially if they are unemployed. Thirdly, the returns are related with the profile of future wages and jobs; several situations may happen:

A worker may profit from general (Becker 1993) or transferable (Stevens, 1996) training relevant to a company, in order to move to a better job. He or she also may seek an agreement with the company regarding future compensation for specific training (Becker 1993, Hashimoto 1981). The returns from LMTPs vary considerably with the evolution of the labour market itself: depressed sector or economies in recession don't reward E&T investments easily. Individual characteristics (namely ability, motivation, work experience) may influence the outcome of the LMTP experience; and quite interestingly, private LMTPs tend to be directed to the "good" (and looked and reliable) workers whereas

public LMTPs tend to enrol the disadvantaged, like the long run unemployed, school drop-outs and the poor.

LMTPs may also support other types of individuals, namely job seekers and migrants. In the case of jobseekers, the Cost Benefit Analysis equates the cost of being present at the counselling session, and the benefit of having a job in the future. The individual may be offered a small job of social nature, or even a job in a private institution; in these cases, generally, the entity is subsidized by the Government at a significant rate (50%) at least during the first 12 months.

In the case of migrants, LMTPs may be essential to guaranteeing support in language skills, administrative legal guidance, job search assistance, and the recognition of diplomas. Migrations being dramatic and painful human processes, LMTPs may facilitate the individuals' transition between two different societies.

Finally, the participation in LMTPs is also influenced by individual characteristics and by societal characteristics: motivated, ambitious and talented individuals, in booming sectors and expanding and affluent economies tend to participate more and to be offered more opportunities of participation by companies and organizations than the average worker. This question is known in the literature as "selection bias". On the contrary, and as a consequence of the social need for equal opportunities, public sponsored LMTPs tend to enrol the disadvantaged, like long run unemployed, poor, adult women, school drop outs, disabled and the law offenders.

#### *Companies and Organizations*

LMTPs may be useful for companies by a number of reasons. E&T are important for companies because they may increase their

output, their productivity, quality product, social climate, and employment. Therefore a company may want to produce or subsidize its own LMTP. Or a company may try to find State produced or funded LMTPs.

An important distinction exists between general training, specific training, and transferable training (Becker, 1993; Stevens, 1996). In the case of general training the company wants the worker to fund the course, because he may leave after completing it; in the case of strictly specific training the company is willing to pay for the program, because the worker can't profit from what he learns; in the case of transferable training, poaching by other companies makes possible the existence of a market failure.

Organizational factors may play an important role defining the position of an organization towards E&T and therefore towards LMTPs. On one hand, Taylorist companies tend to give LMTPs a comparatively small place, due to their pyramidal power structure: only a few workers need extensive E&T and for the vast majority of them specific and limited skills are enough. On the other hand, post-Taylorist companies have a total different approach, and each one of their workers needs a much higher amount of E&T; therefore, for those companies, LMTPs are much more needed.

Some general rules may be used to understand the position of companies and organizations towards LMTPs and E&T. They tend to make contracts with workers in order to profit from their own investment on those workers. They tend to invest more in workers they see as reliable and productive. They also need LMTPs regardless their ultimate goal, their economic sector or their ownership form. Companies and organizations may benefit from support to

migrants, when the migrants are seen as employable workers. Companies and organizations may also benefit from the Government support to contracts, because their costs decrease.

### *Government and Public Authorities*

Market failures may explain the presence of public authorities in LMTPs. But the intervention of the government should only be made, if the potential government failure that is generated by the public intervention is smaller than the market failure that was to be corrected.

Government failures may be of several kinds. They include information failures, when the program is not suited to the real needs; control and evaluation failures, when the program generates much higher spending than it was expected, with no perceived outcome; bureaucratic self interest, which tends to increase the program's expenses to his own benefit; and political interests, linked with the elections cycle, which tends to provide training to potentially influenced voters when elections approach.

There are quite a few market failures that a LMTP may want to correct. There are positive externalities which may be generated from the LMTPs programs, because the knowledge acquired in LMTPs is easily transmitted, and unemployment may be reduced; therefore the social optimum is higher than the private optimum, and public funding is due (Adams, Middleton and Ziderman, 1992; Middleton, Ziderman and Adams, 1991). sub-provision is generated by the poaching externality, that occurs mainly in transferable training and in oligopsonistic markets (Stevens, 1996). Thirdly there are information failures may separate workers from jobs, and therefore public sponsored information systems may be quite helpful (Dougherty and Tan 1997).

When the needed LMTPs are large enough, public goods considerations regarding their funding may exist and be sufficient to ask for public intervention. Capital markets may prevent serious disadvantaged people accessing to E&T and to LMTPs (Drake, 1986, 17; Parsons, 1990 68). Discrimination and screening procedures by companies may exclude many people from privately funded LMTPs (McConnell and Brue 1989; Paul 1989 pp.72-80; Spence, 1975). Market dualism (penalizing SMEs, which don't have the means to make LMTPs, to profit from knowledge and to secure knowledge workers) (Paul 1989 pp.80-90; Chapman, 1993 pp.55-8; Reynolds Master Mozer, 1986 pp.165-6).

Individual myopia may also lead to underinvestment in training or LMTPs; in this case, LMTPs may be considered as merit goods to companies and individuals (Dougherty and Tan, 1997). Finally, big social questions like social cohesion, and regional unbalances may be addressed through LMTPs; inflation pressures and unemployment and also specially long run unemployment may also be prevented by training and LMTPs;

Another reason for the public presence in LMTP is equity problems. Society may consider that it is unfair that some people and organizations don't have the same opportunities to participate in LMTPs as everybody. Furthermore the public authorities may consider that from the unbalance in participation rates, social problems like economic inequality, social exclusion and crime may derive. Finally, it is crucial to note that the rationale for the type of public intervention is quite different. In one hand, public intervention may result in production of LMTPs if there is a social need to be fulfilled and no private or third

sector entity wants to produce the due program. These are limited circumstances. But, in the other hand, the public intervention may be much widely justifiable in the case of funding: it is only necessary that the need exists and the private or third sector entity as not the means to fund it or to find the funds in the banking sector.

#### *Other Entities, Trainers, Families, Unions.*

Cost Benefit Analysis applies to all those important entities, in the LMTPs system. Benefits that matter are not only perceived in monetary terms: trainers may earn experience, families may learn and union may gain social recognition; union membership may be important to determine the participation in the LMTPs, and this may increase the rate of unionization. Costs are also not only monetary, since non participation can generate social exclusion even to trainers, unions and families.

#### *Society as a Whole*

Economies and societies may be divided in "low skills" and "high skills" ones (Ashton and Green, 1996). The relation of that classification with LMTPs is very interesting and important. In low skilled economies competences are rare and badly remunerated. LMTPs are not common.

Most skills are acquired informally. Self training and non formal education are very important in relative terms. Formal training and formal education levels are extremely low, in absolute and relative terms. This is the case of poor, developing and not developed countries. Wages are low, and employments "bad". A "low job, bad skills" trap exists (Snower, 1996): people tend not to invest in skills because they feel they will not perceive the returns from that investment, and companies tend not to create

highly skilled jobs because they feel they will not find the adequate work force.

In emerging and developing countries, the “vicious cycle” that characterizes a “low skills” equilibrium is broken. In these countries an agreement between the State, the companies and the individuals is somehow reach, and life long education becomes a reality. LMTPs become large, and very important, because it is decisive to upgrade the skills levels of the population: even if a “low skills” economy may profit from low wages to benefit from trade, that strategy is doomed in the long run, due to political and social pressures and to the technical evolution.

LMTPs are an essential feature of developed countries, in which a “virtuous cycle” of investment in knowledge exists, and in which investments in Knowledge are common. The virtuous cycle means that people invest in Knowledge expecting to use that knowledge and companies create knowledge rich vacancies because they expect knowledge workers to exist. Skills and competence upgrading are a constant need in those very competitive nations, and therefore, LMTPs, be them privately or publicly funded, are essential in those societies. Table 1 below (see 4), shows how, in the world, the theoretic ideas exposed in this subsection, are translated into practice.

It is interesting to note that the existence of public provision and funding of LMTPs is closely linked to the existence of Welfare States. Several types of Welfare States exist in the world: Liberal, Social Democratic, Conservative, Socialist, and Latin (Esping Andersen 1990, Deacon 2000, Ferrera and al 2000). Welfare States are basically a feature of developed countries. When a country begins to develop, a Welfare States becomes to be built. The societal emergence of LMTPs’ is therefore a small part of a more

complex social evolution, and is influenced by that evolution.

### *Evaluation of LMTPs.*

The evaluation of LMTPs may be made in microeconomic or in macroeconomic terms. Microeconomically, there are two main ways of evaluating LMTPs: by results or by impact studies (Tomé 2005). Results are description of occurrences, and relate to what was done in the project. Results indicators are usually physical (number of supported people, number of organizations involved) or financial (total cost) in nature; sometimes results relate to accessibility, analysing the share of participants in the potential demand. Quite crucially, results evaluation deals with happenings, costs but not with benefits. In order to determine benefits, impact studies have to be made. Impact studies are of two natures: experimental or non experimental (Burtless and Orr, 1986; Heckman Smith and Lalonde, 1999).

In both cases the impact of the LMTP is calculated using an equation with the following generic formulation:

$$(1) Y = AX + BP + ei$$

In this equation Y represents the program impacts, P represents the participation in the program, an very often is a dummy variable, X are control variables representing observable characteristics of participants that may affect Y, and ei is the error term that represents non observable variables. A and B are coefficients to be estimated. This equation is calculated by using statistical programs like SYSTAT, SPSS, LIMDEP. The availability of the data is however a major problem in LMTPs evaluation procedures (see 6.3).

In experimental studies the assignment to the treatment group (of participants) and to the control group (of non participants) is done randomly. This feature is meant to address the “selection bias” problem (see 3.1). That bias would make the rate of return of the project to be wrongly calculated. But ethical and administrative problems arise in experimental studies. Firstly, it is not easy to prevent anyone for accessing a social benefit for the sake of scientific evaluation, And secondly, administrators fear the unpopularity of impact studies. Therefore experimental studies of LMTPs are very rare in practice. Most of the microeconomic evaluations of LMTP are made by non-experimental studies. In non-experimental studies the problem of selection bias is addressed by defining econometrically a selection rule. A second equation is used, with the following generic shape:

$$(2) \quad P^* = C Z + f i$$

In this second equation  $Z$  are possible determinants of participation, and  $P$  is the endogeneized participation of each individual. This variable  $P^*$  should replace  $P$  in equation 1 and should represent the expected situation of participants and non participants regarding participation having in consideration their characteristics that would determine participation. Combining equations 1 and 2 non-experimental studies should define the impact of programs. The discussion on what kind of studies are best is quite interesting (Heckman and Smith, 1995; Heckman Lalonde and Smith, 1999) and actually the two types of are judged to be complementary.

Macroeconomic studies may also define the impact of LMTPs if LMTPs are considered to be big enough to have an impact in the whole society. In this case,

social input output models, or models of the supply and the demand of the labour market may be used to evaluate the national impact of those programs.

### **Comparative Analysis of Nations, Regions, and Continents**

LMTPs may be funded and organized by the public sector or by the private sector. This section presents an overview of the most important LMTPs that exist in the world.

#### *Public Sector LMTPs*

In the United States, LMTPs exist since the early sixties; programs have been funded through the Manpower Development Act (MDTA), the Comprehensive Employment Training Act (CETA) and the Job Training Partnership Act (JPTA) and School to Work Opportunities Act (SWOP). Support was granted to disadvantaged adults, mainly women, and young drop-outs. On the basis of that support was the idea that selected groups of people should receive public support because the private sector would not help them, and therefore, the public provision of that support was socially important.

Meanwhile in the Nordic Countries (Finland, Sweden, Norway, Denmark), in Central Europe (Germany, Austria) and in the Anglo Saxon countries (UK, Australia) LMTPs were also common. In those countries the support was centred in the unemployed specially the long term unemployed and the young people. In particular, the German Dual apprenticeship system was considered to be a benchmark to the other experiences. In those countries a LMTP experience was considered to be important when it supported, yearly, about 1% of the labour force.

In 1960 the European Economic Community began to fund LMTPs programs

in its Member States, using the European Social Fund (ESF). In the eighties and nineties the ESF became a very important way of providing LMTPs. Beneficiary countries were Ireland, Portugal and Spain. In those countries the percentage of supported trainees in the labour force per year raised to 10%. In the 21<sup>st</sup> century the ESF support was divided between the bigger European Countries (Germany, UK, Italy, France, Spain), and the less developed countries (Eastern Countries and Portugal).

Eastern European countries also had a developed framework of centralized LMTPs, during the Socialist experience. In fact, it is well known that the citizens of those countries are counted among the more skilled in Europe; but with the fall of the regimes, and the openness to the West, those systems also endured a transition process. Those countries are somehow an exception, meaning the competence level of their workers, if fully used, should generate a much higher income level than the actual one.

In Japan and South Korea, the public sector intervention is essentially linked to the establishment of a framework, in which the private sector provides the main share of the operations. In Latin countries, both in South America and in Europe (Portugal and Spain), public participation in LMTPs was not large, traditionally. An labour force elite was created by those policies, but the average level of provision was low. This fact was seriously aggravated by the low levels of general education attainment of the population. In China, the Government has had a very strong domination of LMTPs. The actual scheme was introduced in 1999, basically to new urban workers, rural non-agricultural labourers and migrant workers (Bentchermer and Islam, 2001). In India, Africa, and the low income countries of

Asia, public LMTPs are a rarity. The OECD studies on the percentage of the GDP per capita devoted to Active Labour Market Policies, show the tendency described in this subsection.

#### *Private Sector LMTPs.*

Companies and organizations fund and provide LMTPs when they envisage they will profit from those operations. Therefore, the existence of private LMTPs is closely linked with the level of economic development. In the USA, the UK, Canada, Japan, the Nordic countries, Central Europe, and Australia, companies usually use LMTPs to enhance competences, and to gain competitiveness. The OECD statistics on the labour force participation in company sponsored continuing LMTPs clearly evidences that tendency (OECD 2005). Latin Countries such as Portugal, Spain, Italy (particularly the South) and Mexico have traditionally much lower levels of participation in LMTPs. The rest of the world has even lower rates.

#### **Global or Multinational Significance**

The global and multinational spread of LMTPs is considerable. Firstly, LMTP have a significant global impact, because they are at the core of the distribution of income in the planet. Rich countries are those that have large LMTPs, poor countries those that don't. This one, though, is only the "first part of the story", and the static one. However, and secondly, LMTPs also influence decisively the dynamic of countries, because they influence their growth rate and their competitiveness. Countries that have large LMTPs, can rely on knowledge to achieve growth and competitiveness. Countries which don't have large LMTPs have to rely on very low input costs and large scale economies; the first

feature has bad social consequences (low standards of living), and the second one may not exist in many countries.

Thirdly, as the deslocalization process proves, LMTPs have very strong multinational implications. LMTPs prevent deslocalization, because companies have less incentives to leave. They attract foreign investment, because companies are more incentives to come. Finally, LMTPs are probably the best way to counter the unemployment derived from delocalization.

Fourth and finally, as Intellectual Capital and Human Capital are essential to economic development, and as some countries need resources to invest in those programs, LMTPs may be at the core of deep economic integration experiences in the future, by which “central rich countries” will help “peripheral poor countries” to develop with mutual gains (Tomé 2004).

### **Critical Analysis**

To define the limits of the efficient and equitable use of funds devoted to LMTPs is in itself one of the most important questions that the study of LMTP faces. Even if the theoretical basis that justifies the existence of LMTPs is strong and if LMTPs exist in a growing number of countries, as a need to make the development process stable some questions remain, on the usefulness of LMTPs. Those questions are usually addressed in evaluation studies, and are the following: are LMTPs worth, and therefore, should they continue to be made? Are there better ways than others of making LMTPs? In addition, and due to the fact that evaluation procedures are fundamental to criticize correctly the LMTPs, it is also important to analyse the political and administrative conditions in which LMTPs are generated.

### *Are LMTPs Important?*

An important number of evaluations on LMTPs have been made in the last three decades, which have been summarized in recent publications devoted to the subject (Dar and Tzannatos, 1999; EC 1997; Heckman, Smith, Lalonde, 1999; Martin and Grubb, 2001; OECD, 1998). The broad picture that arises from those studies, is that LMTPs have positive outcomes. However, many qualifications must be put to this idea: the benefit from LMTPs in the case of employment probability is much higher and consistent than in terms of wages; and direct benefits usually decrease when indirect benefits like displacement effects and substitution effects are counted.

LMTPs may help the economy and the society, but their success is also dependent from the overall evolution of the economy; in times of crisis the outcomes of LMTPs tend to be small. LMTPs cannot by themselves solve unemployment crisis; they are necessary but not sufficient; their success relies on the overall evolution of the society.

### *Some LMTPs Better than Others?*

Strictly designed LMTPs, aimed at very disadvantaged target groups, responding to well perceived needs, and occurring in good economic environments, tend to have better evaluation results. Wide social participation of companies, unions and other relevant social actors in the management of programs is also beneficial. Very large LMTPs, tend to be less effective than small ones, in economic aspects. But in social terms, large LMTPs may be justifiable; the debate is still open on this question.

### *Political and Administrative Considerations*

In order to make scientific evaluations on LMTPs, important quantity of data are needed. The collection of those data and its



scientific analysis implies a collaboration between the State and the scientific community. However, some problems may arise, on the determination of the evaluation procedures. Gramlich (1990) suggest a model of evaluating the evaluations. In this model the benefits derived from the evaluations are the costs that would be saved if the program stops. The costs of the evaluations are direct costs and the benefits that cease to exist if the program ends. The benefit derived from the evaluation may be also related with three factors: a) the probability that the evaluation was well done; b) the probability that the government takes note of the result of the evaluation; c) the dimension of the program. Therefore, this model recommends that longitudinal data on participants should be collected, and results methods should be complemented with impact studies; evaluations should not be done if the preconceived idea of the administrator is so strong, that it will not be changed by any evaluation result big programs should be more analysed than short ones.

## **Conclusions**

LMTPs are organized structures that provide support to companies and individuals in the labour market. They are usually composed by operations regarding training, support to migrants, information on the labour market and direct support to the employment.

The economic theory provides scientific logic to the existence of LMTPs, because individuals, organizations, and societies, may have benefits from LMTPs that outweigh clearly the costs. The public participation in LMTPs is justified in equity grounds and carefully measuring the relation between market failures and government failures in LMTPs.

In practice, LMTPs are essentially a feature of the developed world. Private LMTPs exist where capitalism is developed. Public LMTPs exist where Welfare States are developed. Both capitalism and Welfare States are a feature of developed and rich countries. When a country begins to develop, both private and public LMTPs get larger because companies invest more and the government has more funds to build public policies. The Eastern European countries are an exception, because during the Socialist years investments in Human Capital were high, but centralized economic planning originated that the people's competences were not fully used.

LMTPs must be evaluated in scientific ways, in order to be studied with rigour. LMTPs alone can't eliminate unemployment; but they may help preventing it and reducing it. Furthermore, small, precise and well designed LMTPs, that answer to social and economic needs can be very useful.

## **Further Research and Governance Innovations**

The Governability of LMTPs is still a very open question. Much of the scientific research in the field aims to justify the money that is spent, and to find better ways of managing and implementing LMTPs. Transferability of best practices is a very complicated matter, because the underlying social, economic, political and legislative conditions may differ, and may influence decisively the program impact.

## **Internet Sites**

USA Department of Labour [www.dol.gov](http://www.dol.gov)  
EU Employment, Social Affairs & Equal Opportunities. <http://ec.europa.eu/social/home.jsp?langId=en.html>

## Selected References

- Adams, A.; J. Middleton and A. Ziderman. (1992) "Planifier la main-d'oeuvre d'après les "signaux" émis par le marché de l'emploi", *Revue Internationale du Travail*, Number 3, pp. 277-298.
- Ashton D. and F. Green. (1996) *Education, Training and the Global Economy*. Cheltenham, UK: Edward Elgar.
- Becker, G. (1993) *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago: University of Chicago Press.
- Betcherman, G. and R. Islam. (2001) (Editors) *East Asian Labor Markets and the Economic Crisis; Impacts, Responses and Lessons*. Washington DC: World Bank/ILO.
- Black S. and L. Lynch. (2005) *Measuring Organizational Capital in the New Economy*. Discussion Paper. [ftp://repec.iza.org/RePEc/Discussionpaper/dp1524.pdf](http://repec.iza.org/RePEc/Discussionpaper/dp1524.pdf)
- Burtless G. and L. Orr. (1986) "Are Classical Experiments Needed for Manpower Policy? " *Journal of Human Resources*, Volume 21, Number 4, pp. 606-39 .
- Chapman, P. (1993) *The Economics of Training*. Manchester, UK: Harvester.
- Card, D. and D. Sullivan. (1988) "Measuring the Effect of Subsidized Training Programs on Movements in and Out of Employment", *Econometrica*, Volume 56, pp. 497-530.
- Dar A., Tzannatos Z. (1999) *Active Labor Market Programs: A Review of the Evidence from Evaluations*. The World Bank
- Deacon B. (2000) "Eastern European Welfare States: The Impact of the Politics of Globalization", *Journal of Social European Policy*, Volume 10, Number 2, pp. 146-61.
- Dougherty C. and J. Tan. (1997) "Financing Training: Issues and Options", *International Journal of Manpower*, Volume 34, pp. 29-62
- Drake, K. (1986) *New Ways in Continuous Vocational Training*. Berlin: CEDEFOP.
- Edvinsson, L. and M. Malone. (1997) *Intellectual Capital*. NY: Harper Business.
- EC. (European Community.) (1997) *Labour Market Studies*. EC: Luxembourg.
- Esping Andersen G. (1990) *The Three Worlds of Welfare Capitalism*. Princeton. Princeton University Press.
- Ferrera, M; A. Hemerijck and M. Rhodes. (2000) *The Future of Social Europe. Recasting Work and Welfare in the New Economy*. Oeiras (Portugal): Celta Editora.
- Hashimoto M. (1981) "Firm .Specific-Human Capital as a Shared Investment", *American Economic Review*, Volume 71, Number 3, pp. 475-482.
- Heckman, J. and J. Smith. (1995) "Assessing the Case of Social Experiments", *Journal of Economic Perspectives*, Volume 4, Number 2 pp. 85-100.
- Heckman, J.; J. Smith and R. Lalonde. (1999) "The Economics and Econometrics of Active Labour Market Programs", in O. Ashenfelter and R. Lalonde (Editors), *Handbook of Labour Economics*, Volume 3A, chapter 31. Amsterdam: North Holland, pp. 1865-2097.
- Kaplan R. and D. Norton. (1994) *The Balance Scorecard*. Cambridge, Mass: Harvard Business School Press.
- Martin J. and D. Grubb. (2001) "What Works and for Whom: A Review of OECD Countries' Experiences with Active Labour Market Policies", *Swedish*

- Economic Policy Review*, Volume 29, Number, pp. 33-49.
- Maurer, H. (1998) *Knowledge Management and Knowledge Transfer: Key Issues of the Information Society*. IICM papers. [www.iicm.tugraz.at/iicm\\_papers/ist98.doc](http://www.iicm.tugraz.at/iicm_papers/ist98.doc)
- McConnell, C. and S. Brue. (1989) *Contemporary Labour Economics*. Singapore: McGraw Hill International Edition.
- Middleton, J.; A. Zideman and A. Adams. (1991) *Vocational and Technical Education and Training*. Policy Paper. Washington DC: World Bank.
- Mincer, Jacob. (1962) "On-the-job Training: Costs, Returns, and Implications", *Journal of Political Economy* Volume LXX, Number 5, Part 2, Supplement, October, pp.50-80.
- Nonaka, I. (1994) A Dynamic Theory of Organizational Knowledge Creation *Organization Science*, Volume 5, Number 1, pp. 14-37
- Parsons, D. (1990) "The Firm's Decision to Train", *Research in Labor Economics*, Volume 2, pp. 53-75.
- Paul J. (1989) «La relation formation—emploi: un défi pour l'économie Economica». Paris.
- Reynolds L.; S. Master and C. Mozer. (1986) *Labour Economics and Labour Relations*. New Jersey: Prentice-Hall.
- Snower D. (1996) "The Low-skill, Bad Job Trap" in Allison Booth and Dennis Snower (Editors), *Acquiring Skills: Market Failures, Their Symptoms and Policies Responses*. Cambridge: Cambridge University Press, pp. 109-24.
- Stevens, M. (1996) "Transferable Training and Poaching Externalities", in Alison Booth and Denis Snower (Editors), *Acquiring Skills: Market Failures, their Symptoms, and Policies Responses*. Cambridge: Cambridge University Press pp. 19-40.
- Sveiby, K-E. (1998) *Measuring Intangibles and Intellectual Capital—An Emerging First Standard*. [www.sveiby.com/articles/EmergingStandard.html](http://www.sveiby.com/articles/EmergingStandard.html)
- Taylor, F. (1964) *Scientific Management*. London: Harper and Row.
- Tomé E. (2004) "Intellectual Capital, Social Policy, Economic Development, and World Evolution" *Journal of Intellectual Capital*, Volume 5, Number 4, pp. 648-665.
- Tomé E. (2005) "Evaluation Methods in Social Policy", *Boletín Informativo de Trabajo Social (Bits)*, *Revista Digital—Escola Universitaria de Trabajo Social de Cuenca*, Universidade de Castilla la Mancha, Numero 8, Setembro 2005.

Eduardo Tomé  
Universidade Lusíada—Porto  
Rua Dr. Lopo de Carvalho 4369-006 Porto  
Portugal  
[eduardo.tome@clix.pt](mailto:eduardo.tome@clix.pt)

### **Institutions and Governance**

Individual choice is taking place inside the existing institutional infrastructure. That structure could change, so economic outcomes could also change without any change in preferences.

Recognizing the importance of institutions is a part of a much broader approach emphasizing a fundamental two-pronged feature of the development process. Economic development is the process of a stable and sustainable increase of the quantities of goods and services. However, any society has to develop not only its productive capacity but also its ability to direct and manage itself. This is what in the contemporary literature becomes more and more frequently referred to as the *issue of governance*.

Governance may be defined as the capacity to control, manage and direct economic and, more broadly, social development while maintaining its sustainability and the political consensus needed to support the continuously changing complexity of directing, guiding and regulatory actions (Madzar 2002:2). The test of the quality of governance and the indicator of its effectiveness is the degree to which it enables efficient utilization and sustainable expansion of available resource. In other words, we evaluate the quality of governance in terms of what it gives through economic growth and how it enhances the welfare of the population. The failure to develop the necessary governance machinery and the accompanying procedural infrastructure could be taken as one of the major determinants of the recently observed slow growth of many economies. Since the

institutional order represents both the operating support and regulatory framework for any kind of governance, one comes to the institutional infrastructure as a strategic determinant of the economic growth and social development.

Differences in available quantities of production factors and unequal access to technology can not account for difference in the development levels and the rates of growth between different countries, but rather what people do with the wealth at their disposal and how they utilize available resources. The basic component to be examined therefore is not the quantity of various goods and factors of production, nor the quantitative relations among them, but the behaviour of people and organizations which produce such factor availability in the first place (Madzar 2002:1). It appeared that the main question is to explain how people manage to achieve considerably differing results out of available quantities of the production factors. During the past few decades, more and more researches have accepted the idea that institutional factors may be one of the major reasons for these differences.

Institutions are the rules of the game established by the law, tradition, moral etc., which forms the framework inside which repeated interactions between individuals take place. They determine the set of choices of individuals by defining both what individuals are prohibited from doing and, in some cases, under what conditions some individuals are permitted to undertake certain activities (North 1990:3-4). Institutions determine available alternatives and thus transaction costs and profitability of economic activities. From the individual point of view institutions produce certain benefits and costs. Benefits are related to the predictability of other peoples behaviour,

while costs are the result of the constraints institutions impose to individual behaviour.

The institutional infrastructure of every society consists of legal, i.e. formal rules (constitution, laws, statutes and other forms of regulative) and informal rules (moral, tradition, habits, business practice etc.). Legal rules establish the political system (administrative structure, human rights etc.), economic system (property rights, contracts) and protection system (judiciary, police). Informal rules are the rules of behaviour established not by laws but by the norms which arise from the sense of duty, responsibility, tradition, religion beliefs, fear, customs, or from all of the above (Pejovich 1998:45). Effective informal rules have thus passed the test of time. It is hard to determine where the influence of law stops and where the influence of informal rules begins and *vice versa*. Sometimes informal rules govern the behaviour of individuals contrary to legal rules, sometimes legal rules facilitate or prevent the application of norms, and sometimes legal rules and informal rules complement each other.

Legal rules affect economic outcomes through the allocation of property rights, while informal rules act through the transfer of information about routine, traditional and desirable economic behaviour. However, legal rules are not only constraints which rational actors must take into consideration in their optimizing calculus. They also enable market exchange and rational behaviour (bounded rationality). Such a role proceeds from the fact that formal rules condition the reservoir of knowledge and information, and thus reduce uncertainty. Prices and markets are not, as it was assumed inside neoclassical orthodoxy, single mechanisms for the transfer of information. Legal rules supply more or less reliable information about likely behaviour of other

people, facilitating decision making and exchange. The importance of legal rules derives from the fact that those rules condition the flows and outcomes of economic activity and introduce an order into economic life which, instead of chaotic disarray becomes arranged whereby the economic transactions take place and stable relationships between economic actors are established according to the existing rules.

We shall start from the assumption that law (legal rules) has economic logic, i.e. that legal institutions (the majority of them) could be explained by the principles of costs and benefits. That law has economic logic is not widely accepted. Many authors consider it as reductionist approach to law. However, the economic approach to law is one way of understanding the legal rules. The critics of the economic analysis of law especially insist on the fact that such an analysis ignores justice which represents the essence of the law. Nevertheless, in the majority of cases there is equality between justice and efficiency. Many situations which could be named by individuals as unjust are also economically inefficient because scarce resources are dissipated. This is the case when someone avoids any responsibility for the compensation of the damage. The concept of justice, however, in the meaning related to economic efficiency could not always and for every single institution be described from the economic point of view. Economic theory is able to determine what one society has to sacrifice in order to realize the noneconomic ideal of justice. Demand for justice is not independent from its price (Posner 1992:27).

An economic approach to law examines whether law promotes efficiency in economy, i.e. considers effects of the changes in legal rules, because these changes alter the distribution of property rights. Since



legal rules shape individual choice and behaviour through incentives and constraints (i.e. influence efficiency of decision-making) we could conclude that efficiency as the main economic concept under the influence of the institutional infrastructure, especially under the influence of legal rules. The activation of private ownership (precisely defined property rights), markets (freedom of contract and free transfer of property rights) and the rule of law (the level of trust in the state i.e. the belief that the state will guarantee and protect property rights and the sanctity of contracts, and that state will be also subjected to law) are the principal presumptions of efficiency.

### **Approaches of Economic Analysis of Law**

There are two possible approaches in the economic analysis of law. Neoinstitutional economics (the best known representatives are Coase, Williamson and North) stresses structural and qualitative questions; for example whether underlying institutions in society stimulate the resolution of economic problems and if there are more efficient institutions from the comparative point of view. This approach examines the connection between economics and law by comparing the economic effects of the different legal rules. Different legal rules produce different economic effects. It is well known that the behaviour of individuals differs under different legal systems. During the 19<sup>th</sup> century, Latin American countries had adopted (with some modification) constitutional and other laws from North America and other economically prosperous countries. However, the results were not like those in North America and other developed countries. The formal rules were the same but the mechanisms of application of these rules, norms of behaviour etc. were not the

same. So, the same formal rules had produced different effects.

Neoinstitutional economics eschews economic transactions that are running smoothly on the basis of information supplied by market prices, and enters into the real world of uncertainty and incomplete knowledge. Neoinstitutional analysis included transaction costs as costs of property rights exchanges. Theories of the firm and consumer behavior are symmetrically treated because it is assumed that economic actors in production and consumption spheres maximize their utility. The maximization principle is applied to all individual decisions and enables comparison of enterprise behaviour in different forms of ownership, as well as behaviour of an enterprise with different organizational forms.

The second approach is based on Pareto efficiency, the main principles of welfare economics. The main difference between neoinstitutional economics and welfare economics is that welfare economics treats institutions (and legal rules) as exogenous variables outside the economic system, upon which the economic system does not have any influence. For an estimate of the effects of changes in legal rules and economic policy (since these changes have redistributive effects) welfare economics applies normative criteria. It endeavours to describe desirable and undesirable situations in the economy, and thus the choice of economic policy measures and legal rules become normative questions. For example, if a formal rule X leads to lower inflation, and rule Y to higher employment, welfare economics introduces normative criteria by determining what is the preferable state. If it is considered that unemployment is a bigger problem than inflation, the application of the rule X will be recommended. On the other

hand, neoinstitutional economics examines which legal rules enable more efficient allocation of resources, regardless of the redistributive effects.

### **Transaction costs**

In order to fulfil their goals, individuals act simultaneously on two levels. The first level is the establishment of institutions, i.e. rules of the game. And secondly, individuals make a decisions on how to use scarce resources for the realization of their goals. On the first level, actors decide about the choice of institutions, and on the second level they make a choice about how to use scarce resources inside the existing institutions. Efficiency in the resource allocation depends upon the chosen legal rules and upon the way in which these resources are used under the constraints imposed by the formal rules.

The establishment of the rules of the game and their use are not cost free. For the adoption of legal rules, their application, preservation and protection, considerable resources are needed. These resources are associated with adopting the legal rules and controlling its application and protection (army, judiciary, police). In other words, these are the costs of establishment and maintenance of the institutional and legal systems. Along with costs linked with the choice of institutions, there are costs associated with the way resources are used: the costs of searching the possible for exchange, costs of negotiating the contract, supervisory costs and costs related to the control of contract realization.

These two groups of costs – those related to the choice of the institutions and those related to the way resources are used – are transaction costs. The main purpose of the institutions, especially legal rules, is to reduce transaction costs and stimulate

economic transactions, i.e. exchange of property rights. Inside the existing institutional infrastructure, a maximization of interests is achieved if the institutions enable the smallest possible level of waste, i.e. reduction of transaction costs.

Changes in laws and other regulative processes affect economic decision-making, i.e. determine whether scarce resources will go to the most efficient use. Every constellation of legal rules has different effects on allocation and distribution of resources and wealth. Also, every change leads to a reallocation of resources i.e. a different distribution of wealth. Tax and privatization laws, for instance affects the transfer of property rights vis-a-vis the distribution of resources and wealth. Economic policy measures also reallocate property rights and lead to redistribution.

The purpose of legal rules is to enable the realization of economic goals, i.e. to support economic efficiency. Since individuals make a choice in a world of incomplete knowledge and uncertainty, the rules of the game enable economic actors to better foresee possible reactions of other actors. This is the reason why stability is a very important attribute of a good legal system.

Legal rules potentially lower transaction costs through the elimination of obstacles which prevent the exchange of property rights. Transaction costs are a synonym for all obstacles to market exchange. These obstacles exist in all economic systems and could be smaller or bigger, pushing resources toward lower or higher efficiency. Legal rules potentially reduce obstacles through the minimization of conflict from divergent interests. Hence, the basic function of law is to protect participants in market exchange from the opportunistic behaviour of contractual parties and to enable the efficient use of resources. Potentially the

costs of the legal mechanism seeking to resolve conflict between divergent interests (contracting parties) are lower than costs in the absence of such legality. For example, law regulates instruments for the security of claims as a potentially efficient means of protection against opportunistic behaviour of contracting parties.

However, we should have in mind that legal regulation is not cost-free. Costs of legal regulation are direct and indirect. Direct costs are the costs of the state activity associated with the adoption and application of legal rules. Indirect costs are related to the possible loss of allocative efficiency if legal regulation does not respect economic logic. Legal regulation should stimulate efficient behaviour of economic actors and ensure the lowest possible level of waste.

From an economic point of view, law is potentially a crucial means of creating institutions as an economic resource. Legal rules are, as other economic resources, scarce goods which provide their services in time. Changes of legal rules alter the existing structure of property rights, and individuals accommodate their behaviour concerning the use of scarce resources.

### **Changes in Legal Rules**

Legal rules could change very quickly, while informal rules typically evolve slowly. Since informal rules provide legitimacy to legal rules, changes in legal rules do not progress as desired if they are inconsistent with the informal rules.

Changes in legal rules could be initiated endogenously or exogenously. Endogenously generated changes represent incremental adaptation of formal rules to changes in the social and economic conditions. The rules of the game emerge spontaneously via repeated voluntary interactions. The role of the state is to 1)

monitor and enforce those rules, and (2) to institutionalize repeated interactions into the law (Pejovich 1998). New legal rules may strengthen interactions between legal rules and informal rules so as to reduce transaction costs.

Changes in legal rules could also be initiated exogenously, when the initiative for a change usually comes from certain interest groups (antitrust laws, labour regulative etc.). Such legal changes may or may not be in conflict with existing informal rules.

Changes of informal rules require relatively long period of time. Through trial and error, individuals continuously develop and evolve rules for mutual cooperation which pass the test of time and are finally institutionalized into tradition and customs. Spontaneous adjustment of legal and informal rules enables low transaction costs, so the economy should be stable and growing. Institutional changes initiated exogenously may or may not be in conflict with the existing informal rules. Considerable resistance may exist to the exogenously generated legal rules. In such cases, additional regulations may be needed for the purpose of accommodating new legal rules to the existing habits, customs, traditions and moral values. It raises transaction costs through additional regulation. Uncertainty and frequent legal changes may thus inhibit the growth of income and wealth.

Informal rules demonstrate incredible resistance, regardless of the intensity of changes in legal rules. From recent history we have an example of the socialist revolution which did not manage to change many informal institutions deeply rooted in the culture of people. On the other hand, more than half a century of centrally-planned regime resulted with the adoption of some habits (reliance on state, protracted



adoption of compromise, undeveloped sense of self-responsibility etc.) that complicate transition to market economy. In the habits and customs of the people in Balkan liberalism and individualism have never been fully accepted; instead, strong inclination towards egalitarianism and collectivism is present. Values that are inherent to socialist society like passivity and apathy, collectivism and deeply rooted belief that society must take care about individuals, subordination to authorities, perception of a market exchange as unfair etc., does not provide a solid ground for the development of market economy. The policy of “leveling” during the socialism produced the atmosphere of indifference, where stability and security became virtues. In such environment, competition, initiative, market entry and exit, growth and liquidation of existing enterprises, have a negative connotation. That caused very strong opposition against the restructuring and systemic changes (new legal rules) at the beginning of transition.

Changes in legal rules alter the existing interaction between legal and informal rules, affecting the level of transaction costs. If new legal rules are compatible with existing informal rules, transaction costs (costs of the protection of legal rules) are lower and society is able to use free resources for the increase of wealth. Otherwise, if new legal rules are confronted with informal rules, transaction costs are higher restricting the possibility for the wealth increase. High level of transaction costs will tend to initiate the changes of those legal rules that are not accepted by the society.

### **Property Rights**

The exchange of goods and services is exchange of property rights. That exchange is realized through contracts.

Property rights represent the widest legal authority over resources under the constraints imposed by the law. They define the relationship among individuals with respect to all scarce goods. Every good represents a set of property rights that inform individuals about the rules of the game inside which they can maximize their interest. It clearly points out that economy and law act in the same job of connecting the individual decisions about the allocation of resources. Lawyers treat property rights as a set of rights which entitle owner the authority to possess, use and dispose of resources. From the point of view of economists, the owner has right to make decisions about the use of resources (to govern and control), to appropriate benefits from their use (to participate in distribution), and to dispose of the two above mentioned rights or of all three (to participate in exchange). Economists frequently use term property rights instead of ownership because distribution of property rights is very important for the efficient use of resources. On the basis of distribution of property rights economists address different forms of ownership and analyse their effects on economic efficiency (i.e. private ownership, state ownership etc.).

Transferability of ownership is the most important attribute of private ownership, since market exchange takes place through the exchange of property rights. It enables the flow of scarce resources toward the most efficient use. Every decision has economic consequences. It means that the owner is always confronted with risk – that is, bears positive and negative consequences of his own decisions. Transferability enables him to minimize the risk, i.e. to make decisions which dispense the risk. For example, buying the stocks of different companies is the way to minimize the risk.

Precisely defined and protected property rights create stimulus for the efficient use of resources. When property rights are not precisely defined (as with the collective forms of ownership), incentives for the exchange are lower because it is not clear what rights are at disposal to the contracting parties, neither could these parties fulfil their obligations. Production is lower since people are not sure whether they will acquire the benefits of their investments.

There is a clear interdependence between property rights and scarcity of resources. Precisely defined and transferable property rights push resources toward the most efficient use. Private ownership most efficiently links the system of punishment and rewards with the decision-making process, providing the incentives for the efficient use of scarce resources (Demsetz 1967).

### **Coase Theorem of Neutrality**

The Coase theorem reads as follows: *The efficient allocation of resources is independent from the initial distribution of property rights if those rights are precisely defined, and if transaction costs are zero.* Individuals exchange property rights over resources in order to maximize their utility. Through market transactions resources change their owners and move toward the most efficient use.

Legal rules determine initial distribution of property rights. In the absence of transaction costs, any initial distribution of property rights will eventually lead to the efficient allocation of resources. Regardless of who holds the rights over resources, exchange will enable property rights and resources to eventually accrue to those who will utilize them in the most efficient way. Outcome of the exchange is always the same: provided that transaction costs are

zero, efficient allocation of resources is established. However, distribution of costs and benefits is not independent from the initial distribution of property rights.

Nevertheless, transaction costs exist, meaning that market does not always provide efficient use of resources. Sometimes the level of transaction costs prevent the exchange and flow of resources to the most efficient use. Existence of market failures (situations where free market does not provide optimal allocation of resources) serve as an excuse for state intervention. State is often considered as an arbiter who intervenes through taxes, subventions, licences, prohibitions, etc. However, state could only accidentally achieve conditions of Pareto efficiency. In addition, through the choice of legal rules (in order to regulate market failures) state always produces some allocative and redistributive consequences - costs for some groups and benefits for others. Hence, although market in some cases fails to provide optimal allocation it does not mean that state regulation automatically will be more efficient.

### **Restrictions to Ownership and Exchange**

State interventions in the economy impose constraints to the exchange of property rights, which protract the exchange and raise the level of transaction costs. It does not mean, however, that these constraints are always inefficient. Some constraints to private ownership are efficient because transaction costs would be higher in their absence. An illustrative example is monopoly behaviour, external effects, etc. As a general rule, we could say that transaction costs increase and economic efficiency decrease, with the multiplication of constraints.

The most frequent constraints of private ownership and exchange are: price control, rationing, obstacles to entry, constraints in labour market etc. If government decide to maximize the price of a certain commodity, issue coupons which ensure certain amount of goods to the bearer, or impose any other constraint, it weakens some property rights. Limitations of the sanctity of property rights reduce the number of market transactions, while limitations of the right of disposal (price control) disarrange allocative function of prices.

When government maximizes the price of a commodity it weakens the right of possession by interfering with the right of the owner to make decisions about his ownership (sanctity of property rights), or his authority to transfer that right to other individuals according to previously negotiated terms (transferability of property rights). Frequently used excuse for the price control is that more people could afford themselves to buy and consume certain goods under the lower prices. However, since price is established below the level which would be achieved on free market, excess demand and shortages exist. Similar situation occurs when government issues coupons in order to enable some social groups to obtain appropriate amount of certain goods under the lower prices. As with the maximization of prices, rationing causes shortages. Due to the price controls, some exchange opportunities that would push resources from lower to higher-valued uses are left unexploited. Through rationing state interferes with the owners right to decide about his possession and with his authority to transfer that right to other individuals according to freely arranged terms. Maximization of prices and rationing cause income transfer from sellers to buyers.

Introducing the barriers to entry also restrain the property rights. Mostly cited reasons for such protection include: protecting workers from cheap imports, safeguarding of public interest, encouraging innovations etc. Whatever the reason for restricting the potential competition, the effect is always the same – restrictions of the owner's set of choices.

The purpose of the determination of minimal wage is to provide income which would enable decent life for workers. However, minimal wage affects the worker's right to use those exchange possibilities which lie under the minimal wage, and affects welfare of those individuals whose productivity is below the minimal wage.

### **Economic Effects of Legal Rules**

Resources are allocated through individual decisions, under constraints imposed by the legal and other rules such as customs, morals etc. Different legal rules result in different economic and allocative efficiency.

Pareto efficiency is accepted as a neutral criterion for the evaluation of the welfare effects of changes in legal rules. According to the concept of Pareto efficiency, existing allocation system is optimal if different allocation will not increase production of any commodity or improve the position of any individual, without reducing the production of other commodities or worsening the position of other individuals regarding the distribution of resources and wealth. If it is possible to change the allocation of resources and improve the position of some individuals without worsening the position of other, these situations are called Pareto improvements. Criterion of Pareto efficiency enables comparison of two states—one before and one after the change. It makes possible to measure effects of public decisions such as

decisions about the change of legal rules, politics etc.

Pareto criterion of optimality is applicable solely to the voluntary transactions. Only then individuals transfer resources (wealth and income) so that every participant enjoys more benefits than before. Conversely, no one would initiate transaction. As far as changes in legal rules are concerned, applicability of Pareto criterion is related to these situations in which consensus about desirable changes has been achieved between all members of a society. If consensus about these changes can not be achieved, there is strong inclination towards *status quo*.

Since changes in legal rules alter the costs and benefits of economic transactions, it is very hard to reach consensus. Every change of legal rules improve position of some and deteriorate position of the others. This is why Pareto criterion is limited to these situations where public choice (choice of legal rules, economic policy measures etc.) does not deteriorate anyone's position. Since every change deteriorates someone's position, Kaldor-Hicks compensation test is used in order to estimate the effects of new legal rules (Kaldor 1939). Compensation test is also met under the term Pareto-Wicksell criterion.

Compensation test examines whether those who suffer loss from the changes in legal rules could be compensated by those who benefit from the changes, provided that the benefits of the winners after they compensate the losers is bigger than their benefits before changes. It is not necessary for compensation to actually take place, it is sufficient that increased benefits of some individuals are enough to compensate the loss of the others. According to Kaldor-Hicks compensation test, state A would be preferable in relation to state B (changes in

legal rules, economic policy measures, political decisions, external effects etc.) if those who win from the transition to state A could compensate those who lose, under the condition that the winners are in a better position than they were in state B.

Fundamental question of any choice therefore is whether that choice leads to the efficient allocation of resources and to the distribution of resources and wealth that will be socially and ethically desirable. Economic theory can not provide an answer to the second part of the question. It can provide an answer only to the question of whether the new legal rules will lead to the efficient allocation of resources (not whether new allocation of resources and wealth are socially desirable).

### Websites:

FindLaw. <http://lawecon.lp.findlaw.com/>

American Law and Economics Association.  
[www.amlecon.org/](http://www.amlecon.org/)

European Association of Law and Economics.

[www.hg.org/redir.asp?url=http://www.findlaw.com/lawecon/](http://www.hg.org/redir.asp?url=http://www.findlaw.com/lawecon/)

### Selected References

Coase, R. (1960) "The Problem of Social Cost", *Journal of Law and Economics*, Volume 3, October, pp.1-44.

Boudewijn Bouckaert and Gerrit De Geest. (Editors) *Encyclopedia of Law and Economics*. Cheltenham: Edward Elgar, 2000,

Demsetz, H. (1967) "Toward the Theory of Property Rights", *American Economic Review*, Volume 57, pp.347-358.

Dnes, A. (1996) *Economics of Law*. London: Sweet and Maxwell.

Hauwe van den, L. (2005) *Public Choice, Constitutional Political Economy and Law and Economics*, Economic WPA.

[www.hg.org/redir.asp?url=http://econwp.a.wustl.edu/months/le](http://www.hg.org/redir.asp?url=http://econwp.a.wustl.edu/months/le)

Hirsch, W. (1979) *Law and Economics: An Introductory Analysis*. San Diego: Academic Press.

Kaldor, N. (1939) "Welfare Propositions in Economics", *Economic Journal*, Volume 49, Number 195, pp. 34-44.

Madžar, Lj. (2002) *Institutional Constraints on South-Eastern European Development —With Special Reference to Knowledge*. Working Paper No.96, Viena: Vienna Institute for Comparative Economic Studies.

Mercuro, N. and R. Tim. (1984) *Law, Economics and Public Policy*. Greenwood: JAI Press.

Miceli, T. (1996) *Economics of the Law: Torts, Contracts, Property, Litigation*. Oxford: Oxford University Press.

North, D.C. (1990) *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Ogus, A.I. and C.G. Veljanovski. (1984) (Editors) *Readings in the Economics of Law and Regulation*. Oxford: Clarendon Press.

Pejovich, S. (1998) *Economic Analysis of Institutions and Systems*. Second Edition. London: Kluwer Academic Publishers.

Polinsky, A. (1989) *An Introduction to Law and Economics*. Boston, MA: Little, Brown.

Posner, R. A., (1992) *Economic Analysis of Law*. London: Little, Brown and Company.

"Symposium on Law and Economics" (1985) *Columbia Law Review*, Volume 85, 899/1116.

Rasmusen, E. (1995) *The Economics of Agency Law and Contract Formation*. Econ WPA.

[www.hg.org/redir.asp?url=http://econwp.a.wustl.edu/months/le](http://www.hg.org/redir.asp?url=http://econwp.a.wustl.edu/months/le)

Veljanovski, C. (1990) *The Economics of Law. An Introductory Text*. London: Institute of Economic Affairs.

Stephen, F. (1988) *The Economics of Law*. Ames: Iowa State University Press.

Tullock, G. (1971) *The Logic of the Law*. New York: Basic Books.

### **Useful Journals:**

*Journal of Law, Economics and Organization*

*Journal of Law and Economics*

*International Review of Law and Economics*

*Journal of Legal Studies*

*European Journal of Law and Economics*

Natasa Golubovic

Faculty of Economics

Trg kralja Aleksandra 11

18000 Niš, Serbia

ngolub@ptt.rs;

natasa.golubovic@eknfak.ni.ac.yu

## Lender of Last Resort

*Chris Moore*

### **Introduction**

It was recognised over two hundred years ago that financial systems need a safety net such as a Lender of Last Resort (LOLR) in order to avoid the ruinous consequences of financial crises. For centuries banking crises have caused enormous welfare losses through wealth destruction, dramatic falls in economic output and job losses. In the Great Depression of the 1930s economic output fell by a third in the U.S. More recently, the 1997 Asian financial crisis cost some countries a significant proportion of GDP. It was a decline in economic output that accompanied banking panics in England during the late 1700s and the 1800s that led the early banking theorists Baring (1797), Thornton (1802) and Bagehot (1873) to advocate that the Bank of England use its unique position in the banking system to assist banks in crisis. Notwithstanding their well-reasoned arguments, the concept of a LOLR initially met strong resistance. However, from the time the Bank of England finally acknowledged its LOLR role in 1866, the LOLR function has been used worldwide to maintain financial stability.

At first blush, the idea of a LOLR safety net seems logical and eminently sensible. It is in nobody's interest to let a banking crisis destroy wealth and jobs. No government will be re-elected if it stands on the sideline while voters' savings and jobs disappear. Nevertheless, the provision of a LOLR safety net is a double-edged sword for its benefits can be ambiguous. The actions of a LOLR may preserve savings and jobs but its provision it is argued, can cause more frequent and severe crises in the future. The problem is that the provision of a safety net,

especially a publicly provided one, gives rise to moral hazard that changes behaviour in the financial sector. Moral hazard is an important public policy question for LOLR. The cost to taxpayers of supplying market liquidity to banks in a crisis, or rescuing a failing bank by injecting capital, or guaranteeing deposits (not a normal LOLR function) can be very substantial.

A study by Hoggarth et al (2005), for example, found that the average cumulative fiscal cost of resolving 33 banking crises in 31 countries ranging from Finland to Venezuela, in the period 1977-2002, was 15% of GDP. An estimate of the average output losses associated with the crises was 13.3% of GDP, the median being (7.1%). In an earlier study, Hoggarth et al (2003) found that higher levels of LOLR support cost taxpayers more, on average, and did more economic damage, than crises with lesser LOLR support. The level of support for banks in the study goes beyond the normal scope a LOLR function, in that it includes deposit guarantee scheme costs etc. but it illustrates the severe economy-wide impact of bank failures and high fiscal cost of support in a financial crisis.

The evolution of the LOLR function continues today despite two hundred years of intellectual and policy development. The public policy issues surrounding the LOLR are many and varied. They range from whether a LOLR is necessary at all, to its optimal design in terms of reducing the trade-off between protecting the banking system and moral hazard. Other important issues include whether LOLR should be a public or private function, and whether the LOLR is naturally a central bank function, or if it more effective to locate it in an independent agency in order to reduce potential conflict of interest. With the globalisation of financial markets it is

argued that the time has come for an international LOLR to deal with the international financial crises. If this is the case then how will an international LOLR compare and work with its national counterparts?

### **Historical Perspective**

Today, LOLR is considered an integral feature of a modern financial system. This has not always been the case. LOLR, like many seminal ideas in economics, gained recognition in the context of the pressure of events (O'Brien 2003). For LOLR, the significant event was the financial instability that accompanied England's conflict with France during the 1790s. Prior to the 1790s the Bank of England (the Bank) had assisted troubled banks. However, in 1793 the Bank declined to act as LOLR and its inaction triggered a financial crisis. The Bank's reluctance to supply liquidity to the country banks forced the government to issue exchequer bills in order to restore liquidity to the bank system. The crisis was the genesis for the writings of Francis Baring (1797) and Henry Thornton (1802) that introduced the concept of a LOLR.

Baring was first to point out that in the banking crisis of 1793, the Bank failed to use its unique position to resolve the crisis and that it should have acted as a LOLR (O'Brien 2003). However, Henry Thornton's writings are considered more significant. He was the first to recognise that LOLR is consistent with monetary policy (control) and that LOLR deals with temporary and self-correcting departures from optimal monetary growth. On the other hand, he noted that failure to act as LOLR will lead to monetary collapse with dire consequences for the real economy. Thornton also recognised that LOLR was a commitment to the banking system and not individual banks,

and that indiscriminate support for an insolvent banks risked moral hazard. Further, he identified that the Bank of England (central bank) was different from other banks in that it held the final reserve, hence had the power and responsibility to lend freely in a crisis.

By 1810 the basic concepts of LOLR were in place. Over the next six decades, a series of financial crises starting with a historically significant crisis in 1825, and followed by others in 1836-37 and 39, 1847, 1857 and 1866, spurred significant conceptual and practical developments in LOLR. It was the failure of the Overend Gurney Company in 1866 that finally saw the Bank acknowledge its role as LOLR. During this period Thomas Joplin and Vincent Stuckey made important contributions to the LOLR debate, the former being described as "the unsung hero of the history of the lender-of-last-resort concept" (O'Brien 2003:17). The developments were finally brought together in the seminal thoughts and writings of Walter Bagehot (1873).

Walter Bagehot restated and built on the key concepts of Baring and Thornton in his celebrated *Lombard Street*. Bagehot's made two specific suggestions: first, that emergency lending should be at a penalty interest rate, and second, that the Bank should pre-announce its intention to act as LOLR in order to calm depositors and reduce the likelihood of a crisis. The rationale for Bagehot's penalty interest rate was twofold. First, it encouraged banks to quickly repay emergency loans, which promptly returns the money supply to its pre-crisis level, and second, it reduced moral hazard by making banks pay for the imprudent behaviour that led to the illiquidity problem. Because all banks have to pay the penalty rate it encouraged banks



to monitor each other's behaviour. Both Thornton and Bagehot believed the Bank should lend only to banks with good security based on pre-panic values.

Following Bagehot's dictum 'to lend freely but at a penalty rate' the Bank successfully prevented financial crises in 1878, 1890 and 1914 from turning into systemic banking panics (Bordo 1990). According to Wood (2003) the Bank's uptake of LOLR along Bagehot's lines cannot take all the credit for the absence of banking panics in the ensuing fifty years. Wood suggests that traditional causes of financial crises such as crop failures, funding wars and fragile banks played lesser roles in the latter half of nineteenth century. Of particular importance were improved agricultural practices, a growing international trade in wheat, and the development of trade finance. At the same time a reduction in bank numbers from four hundred to sixty-six played a role, as did improved fiscal management, all of which helped England achieve a period of relative financial stability.

By contrast, the United States was not as successful as England in avoiding bank panics. Bordo (1990) shows that while both countries in the period 1873-1932 experienced similar macroeconomic conditions and banking ratios associated with bank panics (in severe recessions), the incidence of banking panics was quite different. The U.S. had four banking panics while England experienced none, despite crises in 1878, 1890 and 1914. Bordo attributes England's success to its judicious LOLR action during this time. France, Germany, Sweden and Canada also experienced similar macroeconomic pressures to those of the U.S. and England during 1873-1932 but none of the four former countries suffered a banking panic. It

was timely LOLR action that avoided panics.

Prior to the formation of the Federal Reserve System (Fed) in 1914 the U.S. was prone to banking panics due in part to its fragmented unit banking structure. The U.S. suffered eleven banking panics in the antebellum period (before the Civil War, 1861-65). During this time the First and Second Banks of the United States had some central banking powers but failed to stop the panics. Deposit insurance schemes were established by some states, and 1857 the New York Clearing House Association began issuing loan certificates, but these arrangements were not successful in forestalling panics. The U.S. attempted to establish federal control over the banking system with the National Banking Acts of 1863 and 1864 that allowed for the creation of a system of national banks with a charter set by the federal government. The government controlled national banks had higher capital requirements and lending restrictions imposed on them to make them safer than private or state banks.

Nevertheless, the U.S. experienced three major panics in 1873, 1893 and 1907-08 when the clearinghouses and the national banks were too slow to supply emergency liquidity. The panics only ceased when the deposit-currency conversion was suspended. After the formation of the Fed, with its LOLR mandate, the U.S. experienced no panics until 1930. However, in the following three years banking panics accounted for 40% of banks closing their doors and a reduction in the money stock by a third, (Bordo 1990). In an attempt to stem bank failures, Congress established at President Hoover's request, the Reconstruction Finance Corporation (RFC) in 1932. According to Butkiewicz (1995), the RFC was initially successful in slowing the



decline in bank failures in 1932, however, it ultimately failed to stop bank collapses in 1933 due in part to its sterilized lending policy (no increase in monetary base at a time when money and credit expansion was needed) and partisan politics. The federal deposit scheme established in 1934 greatly reduced banking panics with no systemic banking crises in the U.S. since its introduction.

From the end of World War II until the 1970s were years of relative financial stability. But the past thirty years have witnessed a marked increase in bank failures. Hoggarth et al. (2003) cites some 33 systemic banking crises (when all, or nearly all, bank capital in the system is eroded) in the period 1977-2002. Most but not all occurred in developing countries. In Britain, the Bank of England LOLR actions stopped the 1974 Fringe Bank crisis becoming a full-blown banking panic. In the U.S., a number of banks received LOLR assistance when their solvency became questionable. Most notable being among those banks assisted were, Franklin National (1974), First Pennsylvania (1980), Continental Illinois (1984) and many Savings and Loans institutions failed as a result of interest rate volatility, poor management and corruption over the 1970s and 80s. Similar banking rescues and liquidations took place in Canada and Europe during this time.

The Asian banking crisis in the 1990s left banks in South Korea, Indonesia, Malaysia and Thailand with non-performing loans equal to 30% of total assets (Llewellyn, 2002). Caprio and Honohan (1999) put the total bailout cost of 59 banking crashes in developing countries during the pre-Asian crisis period 1976-96 at US\$250 billion or an average cost of over 9% of GDP. Caprio and Honohan cite estimates of the cost of the Asian financial crisis lying between 20-55%

of GDP, while the cost of financial crises in industrial countries over the same period averaged 4% of GDP. They point out that developing economies are more prone to banking crises because of volatile macroeconomic conditions and frequent regime changes including political interference in bank management and regulation. Rogoff (1999) argues that the cost of banking crises cited above raises doubts about the efficacy of the LOLR function. Moral hazard from LOLR and deposit insurance encourages risky lending especially in countries where moral hazard is not held in check by an enforced supervisory regime.

The rise in recent banking and financial crises has stimulated academic and industry interest in both the theoretical and policy aspects of LOLR at a national level, and increasingly, at an international level in an era of global banking and financial markets.

### **Alternative Perspectives on LOLR**

Although the LOLR function has found near universal acceptance and application, philosophical differences exist concerning its scope and modus operandi in a crisis. Bordo (1990) describes four alternative views, one of which, adherents of free banking, sees no need for a LOLR at all.

The Classical View, represented by Thornton (1802) and Bagehot (1873), advocates lending freely at a penalty rate to solvent banks in a crisis; rescuing illiquid but solvent banks; and announcing at the outset of a crisis the intention to lend to sound banks. This view of the LOLR function has found support with monetarist economists in the writings of Friedman and Schwartz (1963), Schwartz (1986) and Meltzer (1986). They advocate the Bagehotian principles of strong leadership in a crisis, especially the need announce

support for solvent banks at the beginning of a crisis. The provision of emergency liquidity is critical. Friedman and Schwartz believe that the devastating bank failures and contraction in U.S. money stock could have been avoided if the Fed had supplied emergency liquidity in 1930 and 1931. Schwartz goes so far to say that had Bagehot rules been applied consistently in the U.S., as in other countries, there would have been no need for deposit insurance. Meltzer (1986) believes that insolvent banks should be allowed to fail otherwise banks will take greater risks.

Goodfriend and King (1988) argue that LOLR action as a facet of monetary policy and therefore believe that OMOs are the appropriate way to give banks additional liquidity or high-powered money in a crisis. Lending to individual banks via the discount window, however, need not impact on the level of reserve currency if it sterilized (offset) appropriately with an OMO. Goodfriend and King see this as equivalent to a private line-of-credit as neither impact on the money stock but both need to be monitored and supervised. They believe that deposit insurance alone cannot maintain financial stability without the presence of a LOLR.

Goodhart (1988, 1987) takes the view that a LOLR should temporarily assist insolvent banks because it is a myth that it is possible to distinguish illiquidity from insolvency in crisis conditions. Goodhart argues that the cost of re-establishing trust and information ties lost when depositors sever links with an insolvent bank help justify such intervention. Solow (1982) takes a similar view because a bank failure, solvent or not, undermines confidence in the banking system. The LOLR should, therefore, assist an insolvent bank to

maintain confidence and minimise contagion.

According to the principles of free banking school, a LOLR function is not required, because banking panics would not occur if banking were free of legal restriction. They believe the need for a LOLR function is due to the central bank's monopoly on the supply of currency. They argue that restrictions on nationwide branch banking in the U.S. (cannot achieve a diversified lending portfolio) and the inability of banks to issue their own currency (cannot supply sufficient cash in a crisis), increases the likelihood of banking panics, Selgin (1989). Free bankers also assert that contagion is less likely to occur because note brokers operating in the secondary market for bank notes, would supply information on bank health and thus help overcome the incomplete information problem that leads to bank runs in fractional-reserve banking system. They also see a role for clearinghouses solving the asymmetric information problem associated with deposit banking.

## **Financial Crises and Lender of Last Resort Function**

### *Underlying Causes*

Bank failures and financial crises are rarely the result of simple, one-dimensional factors like poor management or corruption. Llewellyn (2002) identifies eight common underlying micro- and macroeconomic conditions, which in combination with country-specific factors have triggered banking crises in developed and developing countries during the past twenty years, as summarised in Diagram 1. First, macro-economy volatility, especially dependence on unpredictable short-term capital flows and speculative asset prices. Second,

structural weaknesses in the economy and the financial system. In particular, weaknesses associated with government ownership of industry and banking, including the misallocation of financial resources and poor corporate governance. Third, hazardous banking practices including lower than minimum capital ratios, excessive FX exposures, too rapid a growth in loans collateralised by speculative asset prices, and poor accounting standards and underpricing of risk. Hazardous incentive structures and moral hazard within the financial system is the fourth factor. Moral hazard may promote an expectation of bailouts, and so encourage imprudent behaviour e.g. cosy relationships between banks, government and corporations, that leads to 'sweetheart lending' that is not negotiated at arm's length.

Ineffective supervision of internationally adopted banking regulations is the fifth factor. This includes, bank capital less than mandated, or below a level that reflects the riskiness of the economic environment. Lenient classification of loan quality and insufficient provisioning were other outcomes of ineffective regulation. Excessive exposure to single borrowers and asset classes, such as property were another outcome. The sixth factor is weak monitoring and supervision. Information asymmetries and deposit insurance (if available) give depositors a limited role and little motivation to monitor banks. This role is, therefore, usually undertaken by rating and/or regulatory agencies. However, supervisory failure is a primary factor in bank crises. Of importance here is political expediency and the failure to intensify bank supervision commensurate with economic deregulation and more competitive (risky) financial markets. Weak market discipline is the seventh factor. The reasons for this

include, the lack of accurate information and transparency; limited opportunities to exercise corporate control (contestable management); and little incentive for wider monitoring of banks due to the monopoly role of official agencies. Finally, unsound corporate governance practices are identified as a common underlying cause of financial crises. These include interconnected lending to related firms, weak internal control systems e.g. risk management, ineffectual directors and potentially hazardous corporate structures, especially conglomerates with cross-shareholdings.

Bordo (1990) argues that banking failures originate from internal and external factors. Internal factors include poor management and judgement, and dishonesty. External factors include adverse changes in the overall price level, and more importantly, relative price changes e.g. of land, financial assets such as bonds, shares, and commodities like oil. Significant shifts in relative prices have the potential to render a bank insolvent if price movements greatly alter the value of the bank's asset portfolio.

### *Bank Depositors and Crises*

Banks, as financial intermediaries, borrow short and lend long. This mismatch of asset and liability maturities makes a bank vulnerable to a liquidity squeeze if its depositors panic and queue up to convert deposits into cash. To raise sufficient cash to meet demand, the bank may be forced to 'fire-sale' its quality assets at a substantial discount to pre-crisis prices i.e. at a loss. In the absence of assistance from other banks with surplus cash, or a LOLR, the previously solvent bank is at risk of insolvency and failing as its liabilities (deposits) exceed its assets (loans and capital). Interestingly, the 'first come, first serve' bank service model exacerbates panics through what Diamond

and Dybvig (1983) describe as a co-ordination problem. If a rumour concerning the soundness of a bank leads depositors to believe that they should cash up deposits, then savvy depositors, who are aware of other depositors' intentions to do the same, will attempt to be first to withdraw deposits before the bank runs out of money. Hence, the urge to be first in the queue can trigger a run, even on solvent banks.

Information about the risks a bank takes e.g. the quality of its lending portfolio, is not easily accessible to depositors. Asymmetric information and agency problems keep depositors, especially small depositors, in the dark about risks a bank takes. Depositors cannot effectively judge the risks associated with bank deposits. However, larger depositors, like corporations, are often better informed of the risks. They often have access to better information through in-house or purchased research. Corporations can use this knowledge to negotiate deposit rates commensurate with the bank's riskiness. The only recourse smaller depositors have, if they have doubts about a bank, is to withdraw deposits. In the absence of deposit insurance a poorly managed bank is likely to experience a run when its depositors become aware that savings are at risk. If deposit insurance exists, however, then small depositors tend to be less inclined to make panic withdrawals.

Freixas et. al (1999) cites three information-based reasons why a LOLR is required to resolve a liquidity crisis due to a bank run or a breakdown in interbank lending (an inefficient interbank market). First, incomplete information about the financial status of troubled banks will make banks with surplus funds (surplus banks) hesitant to lend. Second, to be prudent, a surplus bank should diversify its lending, but to do so in other than a small liquidity crisis,

the bank is unlikely to have sufficient funds to assist enough banks to ensure diversification. Third, banks may become reluctant to lend because they are unsure whether they will be able to borrow to address their own liquidity problem should the crisis become systemic. If the interbank market is not efficient then a LOLR is needed to overcome information deficiencies (send a signal by lending to solvent banks), reassure the market, or directly supply liquidity to the market.

### *The Interbank Market and Crises*

Panicking depositors are not the only source of a liquidity crisis. If a bank cannot renew its short-term interbank borrowing or its other liabilities, such as bonds or commercial paper, then it has a potential liquidity problem. Banks raise a considerable amount of short-term liabilities (money) from each other through the interbank market. While a LOLR has a distinct role in a depositor initiated crisis, it is less obvious that a LOLR is required to intervene in a crisis originating in the interbank market. It can be assumed that depositors are unlikely to place funds in troubled banks (even solvent ones) hence the need for LOLR emergency liquidity. If the crisis stems from interbank lending, then banks with surplus funds should be in a position to assist troubled banks. But this assumes an efficient (informed) interbank market where solvent banks receive funding and insolvent do not. If the market is inefficient (not well informed) then surplus banks will not assist illiquid but solvent banks thus a LOLR must step in and assist.

### *Systemic Risks*

A principal argument for the existence of a LOLR is its ability to stop small banking crisis turning into systemic financial crisis

i.e. the mass failure of solvent as well as insolvent banks such that banking system capital is depleted, or nearly depleted. Financial crises can spread quickly from one bank to another by the process of contagion. If for example it is not possible to identify solvent banks from insolvent ones, then depositors will act rationally and withdraw savings from all banks, irrespective of their soundness. A crisis that hits banks with similar lending portfolios is known as an information-based crisis because depositors realise these banks have high exposure to the same borrowers. For instance, banks with a high proportion of their assets loaned to farmers may experience a bank run if the agricultural sector suffers a harsh downturn due to severe drought or floods. If all banks are hit then the crisis is termed pure panic contagion. Historically, banking crises have taken the form of information-based crises with pure panic crises being comparatively rare (Kaufman, 1994).

The costs, or negative externalities associated with systemic financial crises are substantial as Hoggarth et al. (2003) shows in an analysis of 33 systemic banking crises from 1977-2002. On average, the crises lasted 4.3 years and put the median cumulative output loss at 7.1% to 23.1% of annual GDP depending on calculation method. Systemic crises destroy wealth in the form of savings and investments, and jobs as output falls. Crises also impose costs associated with the re-establishment of new banking relationships. Gathering information to establish a new banking relationship is time consuming. Small borrowers may also find it difficult to re-financing a loan. A breakdown in the credit process accompanying bank failures adversely affects economic resource allocation because of the loss of lending and risk pricing expertise.

If circumstances conspire, small bank crises can escalate into big crises if mishandled. However, big bank failures pose a greater risk to financial stability. The doctrine 'too big to fail' implies that a LOLR must rescue a too-big-to-fail bank because its collapse will likely trigger a systemic crisis, or inflict economic damage in excess of its rescue cost. In practice, the LOLR action acts to avoid contagion from a large bank failure but not necessarily guarantee its survival. Rescuing an insolvent bank is controversial because it is contrary to Bagehot principles and it promotes moral hazard. Nevertheless, Goodhart (1988, 1987) believes an insolvent bank should receive temporary LOLR assistance because it is a myth that a distinction can be made between illiquidity and insolvency in a crisis.

The odds of a systemic crisis increase commensurately with higher levels of unsecured interbank lending, because contagion can spread rapidly to banks with a high exposure to a failing bank. The unsecured interbank lending motivates banks to monitor the soundness of borrowing banks. However, moral hazard associated with LOLR can reduce banks' motivation to monitor unless insolvent banks, including large banks, are allowed to fail. Bank settlement arrangements are also a source of potential systemic risk depending upon structure. Large uncollateralised bank exposures are common with deferred net settlement systems. However, central banks have encouraged the adoption of real-time gross settlement systems (RTGS) RTGS enable banks to settle interbank obligations arising from high value transactions, particularly those associated with foreign exchange and securities markets, on request (throughout the day as transactions occur). Before the introduction of RTGS, banks' net positions were usually settled at day's end,

rather than during the day. If a bank failed at some stage in the working day, then banks with a large settlement exposure to the failed bank could be at risk of failing themselves, as they would be owed substantial sums of money from incomplete settlements. RTGS thus reduces settlement risk but it means the central bank carries some risk in providing uncollateralised intra-day liquidity.

### **Lender of Last Resort In Practice**

The action a LOLR takes to support the banking system depends on the nature of the crisis. In the case of a bank run the LOLR would act quickly to supply sufficient market liquidity (reserve currency or high powered money) to meet demand. The LOLR may even supply liquidity to an individual bank if solvent, or in exceptional circumstances, to an insolvent bank under specified conditions that minimise its financial exposure. If the crisis stems from a major bank failure, then the LOLR has a number of options that may need government approval, if taxpayers' money is required to resolve the crisis. The LOLR could rescue the failing bank with an injection of capital or alternatively, sell it to other banks, or liquidate it and ease it out of the financial system.

### *Liquidity Crisis*

Resolving a crisis by increasing the level of reserve currency in the banking system can be interpreted as a short-term focussed monetary policy response (as opposed to its usual long-term focus) designed to stop interest rates spiking in a liquidity squeeze. Linking LOLR and monetary policy in this way has led some to advocate that emergency liquidity is best provided through open market operations (OMOs) i.e. central bank injected liquidity is secured against quality bank assets (Goodfriend and King

1988). Increasing system liquidity via OMOs helps banks with surplus funds to lend to deficit but solvent banks and in the process creditworthy banks are supported and interest rate rises stemmed.

If the interbank market is inefficient (not fully informed) then it can exacerbate a liquidity crisis because participating banks are unwilling to lend to each other since they are unable to distinguish illiquid but solvent banks from insolvent banks. In such circumstances a LOLR may have to support individual banks. Liquidity support to the market increases the supply of reserve currency, but support for individual banks may not, if offset (withdrawn from the market) by the central bank. Because events move rapidly in a crisis it is difficult to tell illiquid but solvent banks from insolvent banks. Fast changing asset prices, for example, may result in a solvent bank (pre-crisis) becoming insolvent during or after a crisis. The LOLR must therefore have a strategy in place that allows it to exit at least cost to itself or the government.

### *Resolution Strategies*

When an individual bank fails LOLR assistance can take the form of liquidity assistance, and/or a capital injection, often in consort with other banks. The use of risk capital is justified in economic and political terms if the cost of saving a bank is less than the cost of it failing. However, there are risks involved, because the LOLR needs to make a quick decision at a time when it is difficult to identify solvent from insolvent banks. If the bank is solvent but illiquid then the risk is lower. But if the LOLR makes an error of judgement and assists an insolvent bank then reputational, as well as financial costs are at stake. Before assisting a failing or failed bank, the LOLR would generally seek prior government approval and support

to use taxpayers' money, as LOLR do not normally have the financial resources to undertake a large bank rescue. A bank rescue could be money-financed (credit funded by printing money) but in an era of price stability, such a money-financed rescue would not be considered best practice because of its inflationary impact. As a discussion of banking crisis resolution in Latin America emphasises, the first principle for effective banking resolution is: "that a society should exert strong political will to make bank restructuring a priority, allocating genuine, noninflationary public funds to the resolution of the crisis" (IDB 2005:68). The lender of last resort's options to recover the investment includes selling the recapitalised and restructured bank off as a going concern.

The options available to a supervisory agency/LOLR to resolve a banking crisis are circumscribed by the jurisdiction's legal framework and by the nature of the banking crisis i.e. individual bank or systemic. In some countries, like the UK, the responsibilities and procedures to be followed in a crisis are codified in a Memorandum of Understanding drawn up between the principal agencies involved in crisis resolution, namely—LOLR, the Central Bank or the supervisory agency, if separate, and/or the Ministry of Finance/Treasury. Hoggarth et al (2003) point out that it is important to have a clear and transparent resolution strategy at hand in order to avoid pressure for forbearance or force a government rescue. Either outcome may end up with the taxpayer footing the bill rather than the bank(s) owner/shareholders.

The legal environment has an important influence on both the options available to a supervisory agency/LOLR in a crisis, and the speed at which an option can be implemented. For example, the LOLR may

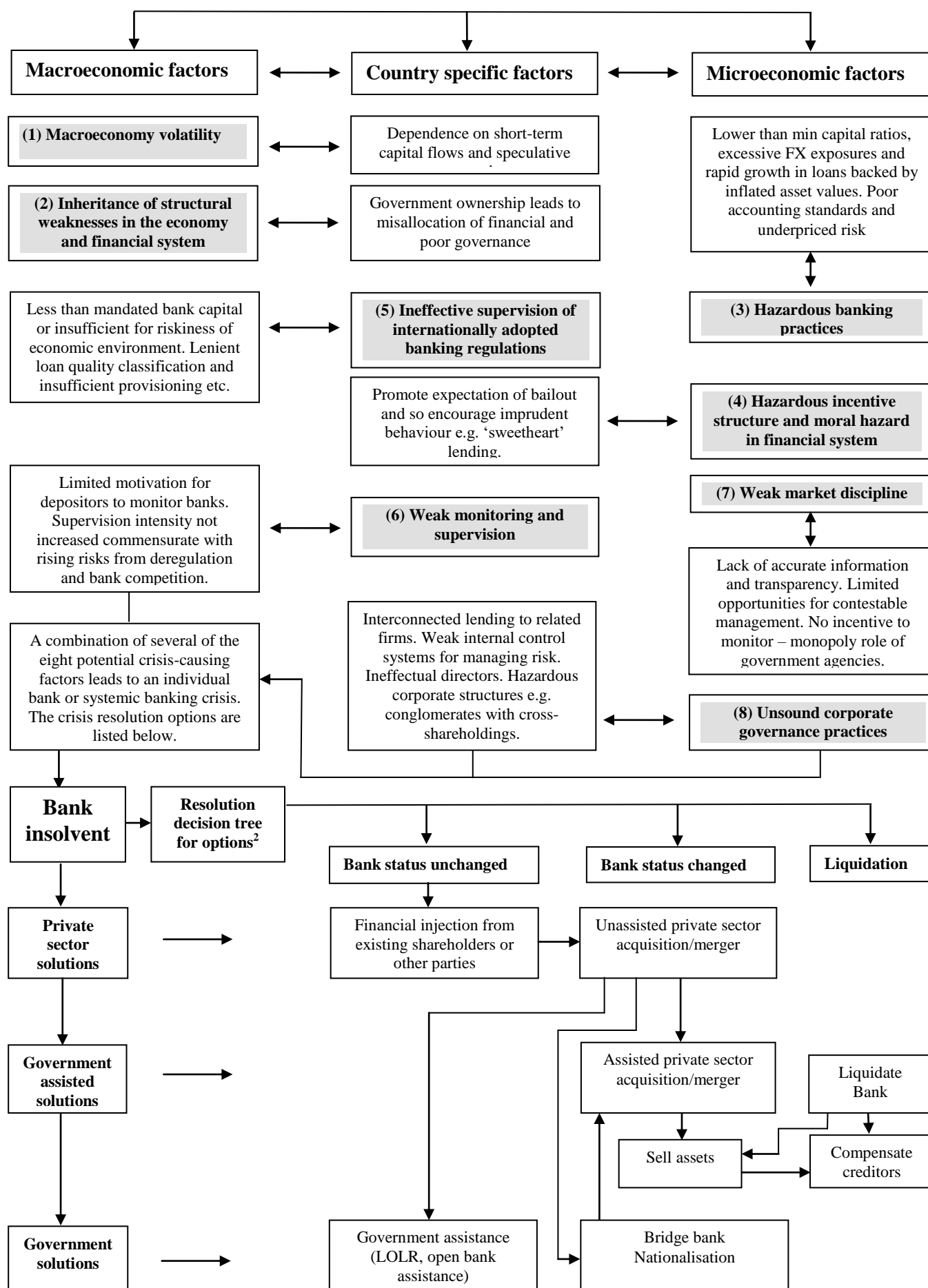
not have the legal right to say, liquidate or sell off a bank. Even if it has the right, it could still find itself open to legal action by bank creditors or owner/shareholders, if its actions at the time are thought unjust or inappropriate. The speed with which bankruptcy proceedings can be filed and actioned also has an influence on the speed of crisis resolution and payouts to creditors. In general, the evidence shows the quicker a bank crisis is resolved, the lower the fiscal and longer term economic cost.

Hoggarth et al (2003) list a number of options available to a supervisory agency to resolve a bank crisis. The options are classified in terms of bank status (unchanged, changed, and liquidated), and source of assistance (private sector, government assisted, and government solution). The array of options is varied. They range from an injection of private sector capital to keep the bank solvent and in business, to liquidation with proceeds paid to creditors. If a crisis is limited to an individual bank or small group of banks, then the preferred solution is for the private sector to supply the necessary capital needed to keep the bank(s) solvent. If this is not viable, or the crisis could potentially turn into a systemic one, then the LOLR would have to decide to either liquidate or provide some government assistance. In an extreme crisis, the government may be forced to take over the bank(s) or underwrite its operation. The range of options are summarised in Diagram 1. They are discussed below.

### ***Unassisted Resolutions***

Unassisted resolutions may or may not involve a change of bank status. The simplest resolution (though none is rarely simple in practice) involves no change of bank status.

**Diagram 1: Factors Causing Financial Crises<sup>1</sup> and Resolution Options<sup>2</sup>**





Once a failed (or about to fail) bank is identified, the supervisory agency/LOLR will move in and determine the severity of the problem. Upon inspection, it may decide to stop some types of business activity or lending, or seek management changes, in order to halt further deterioration in the bank's position. It will then demand that the owner/shareholders put up additional capital, along with other interested investors, so the bank is solvent and able to operate. If the additional capital is not forthcoming the next option is to find an unassisted merger partner, usually another bank(s). The potential merger bank will undertake due diligence, with assistance from the supervisory agency, in order to discover the extent of the problem. In today's competitive banking market no bank would absorb a troubled bank unless the benefits of doing so clearly outweigh the potential cost and risk involved. Success in finding a bank prepared to merge with a troubled bank is likely to be inversely correlated with the size of the failed bank.

### *Liquidation*

If the supervisory agency/LOLR has no success with unassisted options then the next step considered is liquidation. This option declares the troubled bank insolvent, its doors shut and its assets sold off at market prices. The proceeds, net of liquidation costs, are then distributed to depositors/creditors according to their rank in the credit line. In most cases creditors will only get a proportion of their investment back; and that may take many years to achieve. If it is a large bank then the sale of its assets could impact on market prices, hence liquidators have to be prudent in their disposal so as not to depress market prices and adversely affect value of sound banks' assets.

### *Government Assisted Resolutions*

If government assistance is required to resolve a financial crisis there are a number of options available. These can be classified in terms of whether or not they change the status of the troubled bank(s). Interventions that leave bank status unchanged (no takeover of the bank) include temporary LOLR liquidity provision to an illiquid but fundamentally sound bank, or open assistance to a distressed bank, through the purchase of part or all of its non-performing loans or capital injection. These intervention options allow the bank to continue trading, hence give the bank an opportunity to sort itself out, or alternatively, allow time for the supervisor to come up with another option e.g. liquidate or find a buyer. However, open bank assistance is potentially a risky option. If the government does not put conditions on its assistance e.g. change management, or impose some loss of shareholders' equity, then an increase in moral hazard is likely to occur.

Other forms of government assisted interventions change the status of the troubled bank through a change of ownership and control. The interventions include assisted acquisitions and mergers (A&M), or full government ownership. A&M come in different flavours. The failed or failing bank may be acquired or merged wholly or partially (some parts of the bank such as branches or deposits sold off separately) with another bank or financial institution. The resolution authority may assist a merger through a purchase and assumption (P&A) transaction. Here the acquirer purchases the assets and assumes all, or part of, the liabilities but is compensated for doing so. Shareholders' equity is lost in these transactions. Creditors too may lose out, if a partial merger occurs.

A&M resolution does mean it is business as usual for both depositors and borrowers.

An alternative, but temporary resolution strategy to long-term government ownership of a failed or failing bank, is the creation of a bridge bank. Such temporary ownership buys government time to restructure the troubled bank, including tidying up its balance sheet, while working on a final plan to sell it to another bank or financial institution. The final resolution strategy is for the government to take outright ownership of the bank i.e. nationalise it where this is a legal option. This option protects depositors and creditors but wipes out shareholders' equity in the bank. Nationalisation is generally viewed as a second best option to private sector control because government managers do not have the same performance motivation as their private sector counterparts. The government owned banks, on average, perform less well in areas of operating costs and levels of non-performing loans.

Goodhart and Shoenmaker (1995) looked at 24 countries with policies concerning the resolution of bank defaults, and found of 104 failing banks studied, 73 were rescued and 31 liquidated. The study investigated how failing banks were dealt with. Was a rescue package arranged, or were they taken over by commercial bank(s), or put under special administration, or liquidated? It also looked at the source of funds used in the rescue. Did the central bank use its own funds, or did they come from commercial banks, or deposit insurance, or government, or was no external funding used at all? The most popular method of dealing with a failing bank was for them to be taken over by bank(s), often with regulator encouragement. The next most common fate was for a failing to be wound-up.

### *The Issue of Moral Hazard*

Moral hazard is an important and oft-cited problem associated with the LOLR function. Moral hazard occurs when insurance reduces the incentive for the insured to prevent the insured event occurring hence increasing its probability of happening. As the LOLR's safety net is a form of insurance, it creates moral hazard. For example, if a bank believes it will be rescued from its folly, its incentive to behave prudently is reduced, e.g. not putting control systems in place to limit fraud, or encourage quality lending. A well-designed LOLR function, therefore, is one that minimizes both welfare losses and moral hazard. At a general level, moral hazard is mitigated by compliance with bank regulations designed to ensure sound banking practice e.g. minimum capital and liquidity ratios, limits on counterparty exposures etc. The effectiveness of bank regulations and controls in reducing moral hazard depends on the efficacy of bank supervision. Where bank supervision and regulation compliance is weak, moral hazard poses a real threat to financial stability. Goodhart (1988), for example, justifies the central bank as bank supervisor on the grounds that it reduces moral hazard associated with its role as LOLR.

The moral hazard from liquidity provision can be reduced if, as Bagehot recommends, banks pay a premium for liquidity. The premium serves two purposes. First, it compensates the lender of last resort for the risks it incurs in supplying liquidity and second, it penalises banks for mismanaging their liquidity. In practice, however, lenders of last resort often lend at market rates because charging a premium risks triggering a bank run. It may even exacerbate crises by tempting banks to adopt risky 'bet the bank' strategies in order to recover losses. Moral hazard associated with

individual bank rescues is, however, potentially more serious. First, bank management and owners may interpret the LOLR's preparedness to rescue a troubled bank as an opportunity to take excessive risks, or abide imprudent management practices. Second, it reduces the incentives for bank creditors, such as depositors and corporations, and other banks, to monitor and peer review bank performance.

A strategy a LOLR can use to limit moral hazard is for it to employ constructive ambiguity (Corrigan 1990). The objective of constructive ambiguity is to create uncertainty concerning the level, timing and the terms and conditions of LOLR assistance in a crisis. No pre-commitment to a defined course of action, for example, is designed to keep bank management and owners prudent. The assumption is that if banks are unsure of the costs the LOLR will impose in a crisis, then they will avoid taking the excessive risks that lead to liquidity problems or insolvency. To be effective, constructive ambiguity must be handled with skill and aplomb if contagion effects are not to turn a small banking crisis into a systemic one. At times it will be prudent for the LOLR to keep its actions secret, lest it trigger a panic if its actions become public. On the other hand, if a panic has started, then the LOLR's best course of action is to declare its intention to support banks. A declaration of support for banks will reduce uncertainty and restore depositor confidence.

The discretion afforded a LOLR by constructive ambiguity does raise policy issues. The principal ones are the time-consistency problem and a lack of operational transparency. A LOLR can most effectively reduce moral hazard by denying its willingness to assist banks in a crisis. However, it may ex-post (after the fact) find that intervention was the best course of

action. If the LOLR is not transparent in its management of a crisis then it can apply different rules to what appear to be similar situations. Calls have been made for the LOLR to disclose actions taken after the event, as occurred in the case of the Bank of England's management of the 'small banking crisis' in the early 1990s. Constructive ambiguity works best when the bank supervisory regime encourages prudent bank management e.g. mandated minimum capital and liquidity ratios. Better still if management, directors and owners can be held personally liable for imprudent behaviour, as is the case under New Zealand's disclosure-based supervisory regime.

The LOLR can also reduce moral hazard by organising liquidity support through concerted private sector lending (CPSL). If the LOLR is both central bank and bank supervisor, then its superior knowledge of bank financials and its ability to facilitate lending as the bankers' bank, put the LOLR in a position to organise private sector assistance to troubled banks. In this role the LOLR can overcome the co-ordination and information problems that inhibit banks from resolving the crisis themselves. In some cases the central bank may have to apply 'moral suasion' to encourage cooperation because surplus banks (banks with cash) may be reluctant to join the party since they are in a strong position to profit in a crisis. Generally, CPSL is more difficult to achieve in competitive financial markets, hence CPSL has until recently, been practised in the less competitive markets like Italy, France, Germany.

An empirical study by Hoggarth et al (2005) looks at, amongst other things, the relationship between the level of safety net provision for deposits and the likelihood of a banking crisis. They find that an explicit

unlimited deposit schemes increases the likelihood of a banking crisis. The next most likely scheme to increase the chances of a crisis is no scheme (ex-ante) i.e. no official explicit scheme exists but in a crisis the government guarantees deposits. In this situation moral hazard is high even though not explicit safety net exists. The least likely scheme to increase the likelihood of a banking crisis in an explicit scheme with limited depositor protection. Thus a safety net that pre-commits to a restricted cover is most effective at limiting moral hazard.

## Conclusion

LOLR has a long history of intellectual and policy development in its pursuit of financial stability. Pertinent observations may be gleaned from this history. Although bank panics have been rather rare events, LOLR actions have on many occasions been successful in preventing banking panics. When they failed, it was because either the LOLR function did not exist, or if it did, the LOLR took the wrong action - notably, ignoring the basic Bagehot principles of LOLR action. More recently the 1990s have witnessed a number of crises including the severe Asian financial crisis, highlighting that some form of public LOLR function must exist. Past banking crises clearly demonstrate that economic shocks, such as large declines in asset prices can lead to a nationwide panic for reserve currency.

The LOLR function need not necessarily be a central bank responsibility. A government Treasury or finance ministry could carry out the role. Indeed this is a popular model in many countries. However, as Goodhart (1995) points out, a central bank and the LOLR agency must work closely together because the central bank is the only source of immediate LOLR liquidity in a crisis. Deposit insurance is

another answer to the prevention of banking panics. While in the U.S. deposit insurance did stop banking panics, is not a panacea for a panic-free financial system.

Many countries without a deposit insurance scheme have been panic-free. Assistance to insolvent banks, as Goodhart (1987, 1988) advocates, is a relatively recent practice and was rare before the 1970s. Bordo (1990) however, believes that such liberal assistance combined with deposit insurance only encourages risk taking (moral hazard) leading to a greater need for assistance in the future. Recent evidence seems to support this view. Hoggarth et al (2003) show that financial crises in jurisdictions where LOLR support is generous are both the costliest to support and the most economically destructive.

To provide an effective safety net in the future, the LOLR will continue to adapt, as in the past, to the challenges posed by increasingly more sophisticated and borderless financial markets. Goodhart and Illing (2002) highlight a number of public policy issues concerning the LOLR in the context of modern global markets. Is, for example, a LOLR really needed? Free bankers do not think so. What is an appropriate design for a LOLR? To what extent do safety nets encourage moral hazard thus increasing the chances of future financial crises? Should the private sector supply liquidity rather than the public sector? Are Bagehot's principles still relevant today i.e. should liquidity be supplied to the market only via OMOs and not to individual illiquid but solvent banks? Is there a need for an international LOLR given the globalisation of financial markets and how would it work? These are but a few of the questions relevant to the future of the LOLR role. It can be said with some confidence that advances in LOLR theory

and policy will continue as it adapts to future developments in domestic and international financial markets.

### Selected References

- Bagehot, W. (1873) *Lombard Street: A Description of the Money Market*. London: H.S. King.
- Baring, F. (1797) *Observations on the Establishment of the Bank of England and the Paper Circulation of the Country*. New York: Augustus M. Kelly, 1967.
- Bordo, M.D. (1990) "The Lender of Last Resort: Alternative Views and Historical Experience", *Federal Reserve Bank of Richmond Economic Review*, Volume 76, pp. 18-29.
- Butkiewicz, J.L. (1995) "The Impact of a Lender of Last Resort During the Great Depression: The Case of the Reconstruction Finance Corporation", *Explorations in Economic History*, Volume 32, pp. 197-216.
- Caprio, G. and P. Honohan. (1999) "Restoring Banking Stability: Beyond Supervised Capital Requirements", *Journal of Economic Perspectives*, Volume 13, Number 4, pp. 43-64.
- Corrigan, E.G. (1990) "Statement Before U.S. Senate Committee on Banking, Housing and Urban Affairs", Washington DC: US Senate.
- Diamond, D. and P. Dydvig. (1983) "Bank Runs, Deposit Insurance and Liquidity", *Journal of Political Economy*, Volume 91, pp. 401-19.
- Friedman, M. and A.J. Schwartz. (1963) *A Monetary History of the United States*. Princeton: Princeton University Press.
- Freixas, X., C. Giannini, G. Hoggarth and F. Soussa. (1999) "Lender of Last Resort: A Review of the Literature", *Financial Stability Review*, November, pp. 151-167.
- Goodhart, C. (1987) "Why Do Banks Need a Central Bank?", *Oxford Economic Papers*, Number 39, pp. 75-89.
- Goodhart, C. (1988) *The Evolution of Central Banks*. Cambridge, Mass: MIT Press.
- Goodhart, C. and D. Schoenmaker. (1995) "Should the Functions of Monetary Policy and Banking be Separated?", *Oxford Economic Papers*, Volume 47, pp. 539-560.
- Goodhart, C. and G. Illing. (2002) "Introduction" in C. Goodhart and G. Illing, (Editors), *Financial Crises, Contagion, and the Lender of Last Resort: A Reader*. Oxford: Oxford University Press.
- Goodfriend, M and R.G. King. (1988) "Financial Deregulation, Monetary Policy and Central Banking", *Federal Reserve Bank of Richmond Economic Review*, Volume 74, Number 3.
- Hoggarth, G., J. Reidhill and P. Sinclair. (2003) "Resolution of Banking Crises: A Review", *Financial Stability Review*, December.
- Hoggarth, G., P. Jackson, and E. Nier. (2005) "Banking Crises and the Design of Safety Nets", *Journal of Banking and Finance*, Number 29, pp. 143-159.
- IDB. (Inter-American Development Bank.) (2005) *Unlocking Credit: The Quest for Deep and Stable Bank Lending*. [www.iadb.org/res/ipes/2005](http://www.iadb.org/res/ipes/2005)
- Kaufman, G. (1994) "Bank Contagion: A Review of the Theory and Evidence", *Journal of Financial Services Research*, pp. 123-150.
- Llewellyn, D.T. (2002) "An Analysis of the Causes of Recent Banking Crises", *The European Journal of Finance*, Volume 8, pp. 152-175.
- Meltzer, A. (1986) "Financial Failures and Financial Policies", in *Deregulating*

- Financial Service: Public Policy in Flux*, G. Kaufman and R.C. Kormendi, (Editors), Ballinger Publishing Company, Cambridge.
- O'Brien, D.P. (2003) "The Lender-of-Last-Resort Concept in Britain", *History of Political Economy*, Volume 35, Number 1, pp. 1-19.
- Rogoff, K. (1999) "International Institutions for Reducing Global Financial Instability", *Journal of Economic Perspectives*, Volume 13, Number 4, pp. 21-42.
- Schwartz, A.J. (1986) "Real and Pseudo-Financial Crises", in F. Capic and G.E. Wood (Editors), *Financial Crises and World Banking System*. London: MacMillan.
- Selgin, G.A. (1989) "Legal Restrictions, Financial Weakening, and the Lender of Last Resort", *Cato Journal*, Volume 9, Number 2, pp 429-459.
- Solow, R.M. (1982) "On the Lender of Last Resort", in C.P. Kindleberger and J.P. Laffargue (Editors), *Financial Crises: Theory, History and Policy*. Cambridge, UK: Cambridge University Press.
- Thornton, H. (1802) *An Enquiry into the Nature and Effects of Paper Credit of Great Britain*. Edited by F. A. Hayek. New York: Augustus M. Kelley, 1969.
- Wood, J.H. (2003) "Bagehot's Lender of Last Resort: A Hollow Hallowed Tradition", *The Independent Review*, Volume VII, Number 3, 2003.

Chris Moore  
 Department of Finance, Banking & Property  
 Massey University  
 New Zealand  
 C.I.Moore@massey.ac.nz

## Liquidity Trap

*Jan Reijnders*

### **Introduction**

The term liquidity trap was coined by Dennis Robertson, as a situation where conventional monetary policies have become completely ineffective as interest rates are unduly low. If economic agents believe that the rate of interest is so low that the future can only bring a rise in these rates, their expectation is that bond holders would face a capital loss if they stick to these assets. In such a situation, money (that is monetary base or high powered money) and bonds become perfect substitutes. Money demand is perfectly elastic so that it no longer makes sense to inject more high powered money into the economy because this will immediately be absorbed without affecting the rate of interest. Accordingly the monetary authorities will have lost control over the rate of interest (Keynes 1936:207).

The notion of liquidity trap dates back to the Great Depression of the 1930s. It first occurred as a limiting case in Keynes's 'liquidity preference' theory of the demand for money. Due to its potential impact on monetary policy effectiveness, it developed into a major policy issue during the so-called Keynesian revolution after the second-world war. It was underplayed during the inflation years of the 1970s and 1980s, but had a comeback in the recent literature that followed the Japanese slump and the years of low inflation and low nominal interest rates in the late 1990s and late 2000s.

### **Depression Economics**

The liquidity trap is normally associated with an economic slump or depression in which the expected real return on investment is below the real rate of interest. To revive

investment in this situation the monetary authorities would normally set out to lower nominal interest rates which, given the rate of change of prices, would result in a lower real rate of interest. The problem here is that during a depression the expected rate of return on investment may be so low that the nominal interest rate that is required to match it approaches its lower bound, that is the positive rate of interest where the expected long run rate is equal to the risk premium, as in Keynesian case (Keynes 1936; Morgan 1982; Kregel 2004) or where the short run rate comes within reach of its zero bound, as in more recent approaches (cf. Krugman 2000, 1998; Ullersma 2002).

The problem of the lower bound of the nominal rate of interest is further aggravated during a depression when deflation, a general decline of the price level, occurs. In this case the real rate of interest, which equals the nominal rate of interest minus the expected rate of inflation ( $r = i - \pi^e$ ), is higher than the nominal rate. So even if the monetary authorities would bring the nominal rate down to zero, the real rate would remain positive, that is equal to the rate of deflation. With an expected rate of return on investment lower than the real rate of interest, investment will drop. The economy will stall or even start spiralling down and there is nothing the monetary authorities can do to stop it.

According to Hicks such a liquidity trap is one of the characteristic features of Keynesian theory that distinguishes it from Classical economic theory (Hicks 1937: 155). The trap can be illustrated in the context of the IS-LM (originally the IS-LL) model that Hicks utilized to interpret Keynes' ideas. The LM-curve in Figure 1 represents the money market. It consists of the set of all combinations of output and interest rate corresponding to money market

equilibrium at a given money supply. The IS-curve represents the goods market. It consists of the set of all combinations of output and interest rate that correspond to equilibrium in the goods market at a given level of autonomous expenditure. The intersection between IS and LM then identifies the equilibrium levels of output and interest rate that correspond with the joint equilibrium in the goods and money market.

The LM curve is a mirror image of a money market in which the shape of the demand curve is such that demand becomes perfectly elastic at a very low (or zero) rate of interest. This translates into the horizontal stretch to the left of the LM curve. Now suppose that the economy is in equilibrium position A in Figure 1, below, that is characterised by a high level of unemployment:

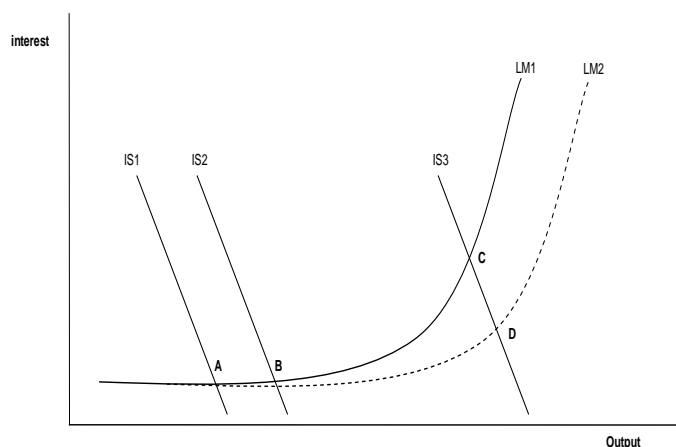


Figure 1: Liquidity Trap in the IS/LM Diagram

If the monetary authority wants to reduce unemployment it will attempt to stimulate the economy by trying to bring down the interest rate through an increase in money supply. An increase in money supply makes the LM-curve shift to the right (LM2) but because the equilibrium position is in the horizontal stretch of the curve, this shift has no effect on either the interest rate or the level of output and employment. The extra

bank money is immediately absorbed into 'idle' cash balances. Therefore it does not lower the interest rate and consequently the rate of output and corresponding employment remain the same (equilibrium output remains in A). Under such circumstances monetary policy becomes impotent. It can not be used to rekindle the economy under conditions of depression.

Under the circumstances fiscal policy is more promising. The government might consider to follow Kahn's and Keynes's advise to increase employment directly by means of public works. This would lead to an increase of autonomous expenditure which makes the IS-curve shift to the right (IS2) and hence directly (and indirectly through multiplication) increases the level of output and employment (point B in Figure 1). Depending on the magnitude of the fiscal stimulus the shift of the IS-curve could be so large that the economy is manoeuvred out of the 'danger zone' of the liquidity trap (to IS3). In this way fiscal policy not only rekindles the economy (the reason for calling it 'pump priming' (Samuelson, 1940)) but at the same time revives monetary policy by creating breathing space for it (a monetary stimulus would in this case move equilibrium from C to D).

The Great Depression and the corresponding liquidity trap were overcome by the preparations for and the waging of the Second World War which, from the point of view of economic theory, can be considered an example of a rather awkward type of fiscal stimulus by 'public works' (Kalecki 1935; Krugman 1998).

### Fiscal Keynesianism and the Demise of the Liquidity Trap

The theoretical and practical possibility of a liquidity trap and the concomitant impotence of monetary policy as compared to the



apparent success of fiscal policy measures had far reaching consequences for economic policy in the age of the Keynesian revolution after World War II. It was thought that the monetary policy instrument (favoured by Keynes himself) had only limited applicability. Under the conditions of a depression monetary policy is of no use because - to draw on an expression uttered by representative Goldsborough when questioning Governor Eccles at the Hearings for the Banking Act of 1935, "You cannot push on a string" (Congress 1935:376; Orphanides 2004:105). Fiscal policy on the contrary was considered superior because it would be as effective during depression as under normal or even booming conditions. This interpretation is for instance implicit in the so-called Radcliffe report commissioned by British Parliament to make recommendations with regard to the effectiveness of monetary policy in the post-war era (Kaldor 1960; Radcliffe 1959). It gave rise to a particular interpretation of Keynes's theory that became dominant in the first decades after the Second World War. This interpretation - that has been given various names such as 'Fiscal Keynesianism', '45-degrees Keynesianism', or 'Hydraulic Keynesianism' (Coddington 1997; Morgan 1982; Patinkin 1990) - puts main emphasis on fiscal policy as a tool for stabilizing the economy.

The period of reconstruction after the war that coincided with the first phase of the Keynesian revolution was characterised by a strong growth of output and employment especially in Europe and Japan (cf. the German and Japanese 'economic miracles'). Interestingly enough this fiscal impulse driven process that fed the extended boom after the war had as a side effect that monetary policy gradually rose again from the ashes of the great depression. The first

reason for this is that the end of depression economics created the elbow room wherein monetary policy could thrive again (see IS3 in Figure 1). The second reason is that fiscal and monetary policy joint forces to produce inflation, an effect that ultimately wiped out all memories of the liquidity trap.

What happened is that fiscal policy became the cornerstone of economic policy. Initially this was supported by a very accommodative monetary policy. In line with the idea of the liquidity trap, money demand was considered to be so elastic that it couldn't exert a direct influence on aggregate demand. In other words it was thought that money doesn't really matter. But it came in handy that keeping money in plentiful supply strongly supported fiscal policy. By lowering the interest rate it reduced the burden of financing the government debt. On top of that it became much easier to service the debt because the resulting inflation reduced its real value (cheap money policy).

Conventional Keynesians were rather cavalier with regard to the dangers of inflation. In line with the idea of the so-called Phillips curve they thought that there was a permanent trade-off between inflation and unemployment. The unemployment rate could be pushed down by stimulating the economy through fiscal impulses (shifts of the IS-curve) supported by an accommodating monetary policy (shifts of the LM-curve). This stimulus would create an upward pressure on the general price level. Accordingly, inflation could be considered the 'price' of low unemployment. It was as if the policy maker had a choice between two options: if you want low unemployment you have to accept higher inflation, and reversely, inflation can be held in check by higher unemployment. In view of the facts that socio-political pressure to

beat unemployment was relatively high and that the potential dangers of inflation were underestimated, Keynesian mainstream economic policy had a built-in inflationary bias (Barro & Gordon 1983; Kydland & Prescott 1977).

It is particularly the focus on the Phillips curve that was targeted by Milton Friedman and the Monetarist in their all out attack on the Keynesian mainstream. In Friedman's view policy impulses will only have a real effect on output and employment in the short run. In the medium run, however, the increase in activity will lead to rising prices. Wages will adjust to this with the consequence that the unemployment rate eventually returns to its initial level. To keep the unemployment rate below par, the authorities have to step up their stimulus and initiate a new round of price rises. The upshot of all this is that an effort to permanently increase the rate of unemployment below a certain level (dubbed the 'natural rate of unemployment') requires to continuously step up the stimulus and hence initiate a process of accelerating inflation (Friedman 1968:1977).

Even on a benign view of the Keynesian bias towards fiscal policy this has important implications for the liquidity trap. Firstly, because the concomitant continuous stream of fiscal impulses gradually shifts the IS-curve to the right. As we have seen earlier this implies that in the course of time the economy is moved out of the horizontal stretch of the LM-curve, i.e. away from the liquidity trap 'danger zone'. Secondly, the Keynesian tendency 'to ride the Phillips-curve' ushers in a period of accelerating inflation. It is precisely this accelerating inflation that makes the liquidity trap evaporate. Since the real interest rate is equal to the difference between the nominal interest rate and the expected rate of

inflation ( $r = i - \pi^e$ ) it is easily understood that the room to manoeuvre for monetary policy increases with the rate of inflation. In other words inflation is an effective antidote to the liquidity trap.

Milton Friedman did not share the sympathetic view on the Keynesian bias towards fiscal policy. He particularly set out to make the case for the opposite view. He particularly wanted to establish that monetary policy is very powerful under all circumstances. In his view the shape of the LM-curve is not at all like the one envisioned by the Keynesians. In terms of Figure 1 this implies that the only relevant part of the LM-curve is not the flat, nearly horizontal, stretch to the left of the curve as the Keynesians would have it. It is rather the steep, nearly vertical, part to the right of the curve that is relevant for the real economy. It is obvious that if reality conforms to Friedman's case, any shift in the LM-curve will have a strong effect on output and employment. Hence monetary policy proves to be very powerful under all circumstances. On this view there is no room for a liquidity trap and the concomitant 'impotence' of monetary policy.

Milton Friedman flatly denied that the liquidity trap had ever existed or would ever come into existence. In his view the monetary situation during the Great Depression of the 1930s was not characterized by a liquidity trap at all. He attributed the depression with its strong downward tendencies and the sluggish recovery to a monetary contraction engineered by the central banks of the time. In his eyes the depression was the manifestation of policy failures rather than of an impasse created by a liquidity trap (Friedman & Schwartz 1963).

Under the influence of the monetarist counterrevolution launched by Friedman and

his allies and the practical experience with several decades wherein the economy operated in an inflationary environment the reminiscence of the 1930s and its main features gradually disappeared. This explains why the liquidity trap that was conspicuous by its presence in the economics textbooks of the 1950s and 1960s gradually sank into oblivion in the 1970s and 1980s (Boianovsky 2004).

### Return of Liquidity Trap

The persistence of inflation created a mood in which the danger of a liquidity trap was thought to have evaporated. This mood, however, changed drastically when Japan, the second largest economy in the world, got into serious trouble. In 1989 the Japanese asset bubble burst which triggered a recession that, in the course of the 1990s, degenerated into an outright economic slump. Japan that was previously set as an example because it sustained such very high growth rates for such a very long time, now had to learn to live with considerably lower growth rates. Real GDP growth dropped from an average of nearly 4% in the 1980s to just over 1% in the 1990s. The steep drop in real growth is accompanied by a substantial rise of the unemployment rate. This suggests that the lower growth rates should not be attributed to a lowering of potential output but rather to an opening of the output gap (the difference between actual and potential output), the manifestation of a slump.

The development of the output gap is represented in Figure 2, together with two other indicators that are relevant to the present discussion, the rate of interest (Libor6m) and the rate of change of the GDP deflator. Although there is substantial discussion about the absolute magnitude of the output gap (Krugman's estimate is 10% where IMF gives 1.2% for the gap), it is

obvious that the gap has opened around the middle of the 1990s and remained there until the middle of the first decade of the new millennium. The existence of the gap coincides with the period in which the GDP deflator becomes negative, that is where deflation begins. At the same time the short run interest rate (Libor) descends below the 1% limit and quickly approaches the zero bound towards the turn of the century. All signs are turning red thus signalling the return of the liquidity trap.

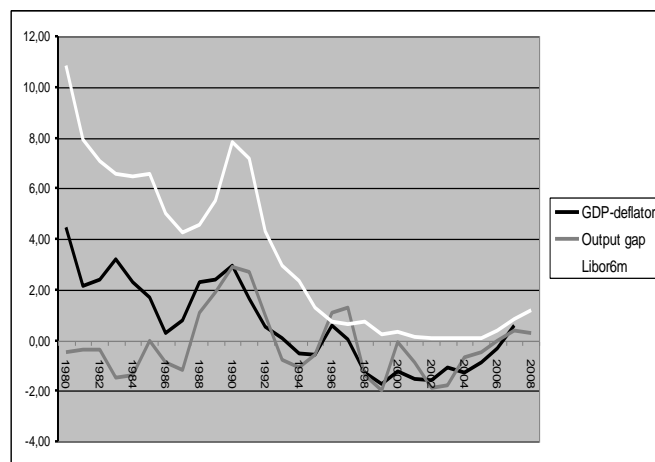


Figure 2: Indicators of Liquidity trap in Japan.

Source: IMF

Paul Krugman was fairly quick in reading the signs of the time. Already in 1998 he suggested that the liquidity trap had returned (Krugman 1998). He set out to reconsider the underlying theory in the light of modern developments in macroeconomic thought. He identified three strands that should be incorporated: first the intertemporal nature of economic decisions and the crucial role of expectations, second the effect of the openness of the economy, and third the role of financial intermediaries. Much to his own surprise Krugman, who originally thought that the liquidity trap was a figment of the limitations of the IS/LM approach, discovered that the liquidity trap survived scrutiny by his intertemporal approach (Krugman 2000:227).

On the basis of a neat two period macroeconomic model he demonstrates that a liquidity trap is conceivable even at the margin of a flex-price economy. A Hicksian type liquidity trap (with real effects) is possible when a ‘Keynesian touch’ in the shape of price stickiness added to the model. The Hicks/Krugman representation of the liquidity trap is depicted in Figure 3.

The line MM is the analogue of the Hicksian LM (LL) curve. In this case it represents a ‘cash in advance’ constraint that is binding as long as the nominal interest rate is positive. The curve CC is the analogue of the Hicksian IS curve. In this case it represents the outcome of rational agent’s intertemporal utility maximization. It illustrates the joint determination of output and the rate of interest. Output is determined by consumers demand, which is decreasing in the rate of interest. Since the curve is drawn for a given expectation regarding the future price level, there is a CC curve for every level of expected future prices. The central bank has the option to determine the supply of money by open market operations. Changes in money supply directly affect the level of the ‘money in advance’ constraint. Accordingly the central bank can move the MM line by monetary policy. Monetary easing shifts the line to the right and hence equilibrium output increases.

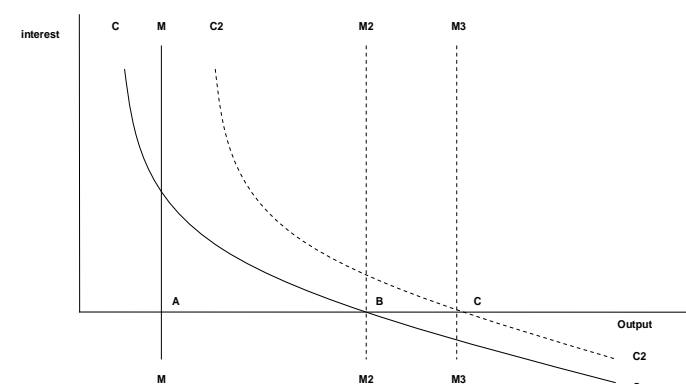


Figure 3: Hicksian Liquidity Trap in Krugman Model

Now let us suppose that full employment obtains with potential output level C. Also suppose that current money supply is such that the economy is in equilibrium at the intersection point that corresponds to an output level of A. Accordingly there is an output gap of A-C. In order to close the output gap the central bank has the possibility to buy bonds so as to increase money supply. This lowers the interest rate thus increasing consumption which in turn increases output and employment: the MM line shifts to the right. The big problem however is that the central bank is not capable of shifting MM beyond intersection point B. At this point the rate of interest will become zero. Any increase in money supply beyond the level that causes the rate of interest to become zero will simply be substituted for bonds, with no effect on consumption and output. Accordingly full employment is one bridge too far for open market operations. Monetary policy has become impotent: The classic Hicksian liquidity trap is back. To his own satisfaction, Krugman has been able to demonstrate “that a fully specified model, fudging neither the role of money nor the necessity of making intertemporal choices, can indeed generate a liquidity trap”(Krugman 1998:150).

In Krugman’s eyes the liquidity trap has returned with a vengeance. Does this mean that under the circumstances monetary policy will again be rendered impotent? The answer is no. It is only temporary monetary expansions that are ineffective. If, however, a monetary expansion is perceived to be permanent, the expected future price level will increase permanently. This will shift the CC-curve to the right (to CC2 in Figure 3). By accommodating this shift by an increase in the money supply (to MM3) full employment equilibrium could be achieved.

All this implies that monetary policy can be effective even in a liquidity trap provided that the monetary authorities can convince the public that its expansionary stance will be kept upright for a considerable period of time. The liquidity trap is thus an expectational issue. It can be overcome by creating the expectation of inflation. In principle, the central bank could generate such an expectation, but the problem is that its promise to create inflation may not be credible. The impotence of monetary policy in a liquidity trap is thus nothing but a looking glass version of the standard credibility problem (Krugman 1998:161).

### **Policy Options**

After the return of the liquidity trap in its Japanese guise and its comeback in economic theory after Krugman's restyling effort, it seems natural that also the well tried Keynesian solutions to the problem are rehabilitated. But things are not quite as easy as that. In the wake of Krugman's reading of the signals emanating from the Japanese economy and his redefinition of the liquidity trap as a credibility problem a lively discussion arose with regard to the direction that economic policy should take. Several options are mentioned.

#### *Inflationary expectations*

In Krugman's view the optimal solution of the liquidity trap problem under the assumption of rational expectations and intertemporal optimisation is the reduction of the real interest rate by creating expected inflation. In principle, monetary policy could be used to achieve this. The main problem is, however, that such a course of action would not be credible. The credibility problem exists because the same Central Bank that formerly put a lot of energy in establishing and maintaining its reputation as

an inflation fighter now suddenly changes to the opposite role and avows to a role of an agent who is committed to higher inflation and makes an effort to "credibly promise to be irresponsible" (Krugman, 2000, 1998). Even if this commitment would be credible in the short run, the public might expect that the Central Bank will revert to its original 'inflation fighter' guise as soon as the liquidity trap is over. For this reason Krugman suggests that the central bank should make a strong commitment to inflation for a long period of time. In the case of Japan it should set a relatively high inflation target of say 4% and keep this for a period of say 15 years to overcome the liquidity trap. The effect could be enhanced by a long run commitment of the central bank to keep the nominal interest rate at zero. This would affect expected future short rates which brings down the long rates and hence stimulates private expenditure.

An alternative to setting an inflation target is to lay down a price level target for a number of years. The advantage of a price level target is that if actual inflation falls short of the target rate in a certain year this must be counterbalanced by above target inflation rates in the future (Bernanke 2002, 2003).

It should be added that all this is not without risk. It goes without saying that the suggested policy of inflation rate or price level targeting must be supported by a sustained expansion of the monetary base. Such an expansion only works to spur inflationary expectations if it is seen as permanent. The big problem then is to determine the speed and the limits of expansion. If it is too small it might not work at all because the central bank's apparent frugality may undermine its credibility. If it is too large it might lead to the necessity of drastically reducing it once

actual inflation starts to accelerate and threatens to develop into a hyperinflation. This also makes the initial commitment to monetary expansion not credible. In Eggertson's view the credibility problem cannot be resolved in this way. Because the Central Bank's actions are discretionary it cannot really commit to future policy. For this reason Krugman's inflation targeting is deemed ineffective unless it is supplemented by other policies (Eggertsson 2003:4)

### *Gesell Money*

Krugman's policy recommendation aims at creating a negative real interest rate by cranking up expected inflation in order to move away from the floor. An alternative policy might be to attempt to lower the floor itself. With given inflationary expectations an alternative route might exist if it were possible to push the nominal interest rate below its zero bound. An unorthodox way of abolishing the zero bound is proposed by Buiter, Panigirtzoglou and Goodfriend who suggest pushing the nominal interest rate below zero by introducing Gesell money. This is achieved by requiring a negative nominal rate of interest on cash holdings through levying a tax on money (Buiter & Panigirtzoglou 1999; Buiter 2001; Goodfriend 2000). The problem with this approach is that it is administratively cumbersome and will encounter major technical problems especially with regard to a tax on currency (Svensson 2003:154).

### *Fiscal Stimulus*

Another plausible strategy is of course the time honoured policy of stimulating the economy out of its liquidity trap by means of fiscal policy. Interestingly enough Krugman is not really in favour of this. There are two reasons. The first is a theoretical one. Krugman's model is built on the

representative agent, intertemporal optimization approach which logically involves adherence to the Ricardian equivalence theorem. This implies that an increase of government expenditure is interpreted by economic agents as an increase in future taxes to which they react by reducing their current expenditure. This neutralizes the effect of the original fiscal stimulus. The second reason has a more practical background. The question is whether a single fiscal impulse can jolt the economy out of the doldrums. In his view this would require a multiple equilibrium situation where the economy got stuck in a slump-equilibrium. The fiscal impulse could then push the economy at once towards a higher level equilibrium. If the multiple equilibrium story doesn't fit, however, there might be no such cheap trick available. Complete recovery might then require a continuation of the stimulus for an extended period of time. It is doubtful whether the consequent accumulation of government debt will be sustainable in the long run (Krugman 1998: 160). Japan's government went this road but it did not meet with great success. Because the fiscal stimulus was financed by increasing government debt that was already high from the outset. This increase in public debt was deemed unsustainable because of the possibility that the private sector anticipates future tax increases and takes measures to cope with this calamity by reducing its current spending (hence the fiscal stimulus crowds out private expenditure in line with the earlier mentioned Ricardian equivalence proposition).

Krugman's position is a matter of controversy. Sumner, for instance, doesn't agree with Krugman's deduction that an increase in government debt will not be sustainable. In his view the burden of

government debt for future generations can be reduced by accommodating the fiscal stimulus by a strong monetary expansion (Sumner, 2002: 489). It therefore appears that monetary policy can still contribute to organizing the escape from the liquidity trap. According to Sumner it is especially the fact that monetary policy is relatively ineffective that points in favour of using monetary expansion to move out of the trap, be it as a support act to ameliorate the effects of a fiscal expansion. Also, Eggertson objects to Krugman's reasoning. In his view, it is precisely fiscal policy together with deficit spending that have to be added to the monetary instruments to be able to overcome the credibility problem. Higher nominal public debt gives the government a strong incentive to inflate in order to reduce the real value of the debt. Provided that the central bank and the treasury coordinate their policies with the aim of maximizing social welfare, deficit spending increases output and prices. Accordingly the central bank has to convince the public that it will accommodate fiscal policy and deficit spending. It was the unwillingness of the Bank of Japan to coordinate with the treasury that explains why the build-up of nominal public debt in Japan so far failed to increase inflation expectations (Eggertsson 2003).

Another case in favour of fiscal policy is made by Kregel. He contests Krugman's diagnosis of Japan's troubles. The problem is not that the real rate of interest is too high because the nominal interest rate cannot go below zero. The problem is rather that due to structural changes the expected real rate of return on investment dropped below the level of the real interest rate. The diagnosis should thus not be that Japan is in a liquidity trap but rather that Japan is in underemployment equilibrium due to

deficient demand. What is required is therefore not a credible inflation policy but rather a policy that credibly raises aggregate demand (Kregel 2004).

### *Pigou Effect*

It goes without saying that the Monetarist position regarding the liquidity trap will considerably differ from the Keynesian stance. Although the modern representatives of the monetarist approach do not go as far as Milton Friedman by categorically denying the possibility of a liquidity trap (they learned from the Japanese experience) they are less convinced that the liquidity trap presents an acute problem. They have in common with Krugman the idea that monetary policy still counts under liquidity trap conditions. They differ from him with respect to their assessment of the gravity of the disease and the degree to which discretionary policy is called for. Apart from Krugman's transmission that works through the interest rate, Monetarists particularly stress the importance of alternative transmission channels, such as relative price adjustments of non-monetary assets that are transmitted along the yield curve and that also affect the exchange rate (Brunner & Meltzer 1988; Meltzer 1995, 2001b, 2001a)) or through the credit channel (Bernanke & Gertler 1995). In a general sense these alternative transmission channels to which Monetarists refer can be subsumed under the heading of the 'Pigou-effect' or 'real balance effect'. According to the Pigou-effect, deflation creates a wealth effect in the goods markets. Decreasing prices increase the real value of money and non-monetary assets. In other words real wealth increases when prices go down. Since real wealth lies at the root of private expenditure, aggregate expenditure will increase. In terms of Figure 1 this implies that the Pigou-effect will

gradually shift the IS-curve to the right. Given time the market mechanism when left to its own will gradually pull the economy out of its liquidity trap. Accordingly there is no need for discretionary monetary policy.

### *Currency Depreciation*

Provided that the liquidity trap occurs in a single country, depreciation of the currency could be contemplated (McCallum 2000, 2002, 2003; Meltzer 2001a; Svensson 2001, 2003). Depreciation is a positive impulse for the economy because it directly stimulates the export- and import-competing sectors and could thus be used to escape from the trap. In view of the connection between monetary policy and the exchange rate, expanding the monetary base (relative to the one of the main trading partners) will anyway involve depreciation of the domestic currency (Meltzer 2001a). But here again the fruits of the expansion can only be reaped if it is seen as permanent. The earlier mentioned credibility problem remains.

To overcome it, Svensson proposes to target the exchange rate directly. The monetary authorities (depending on the institutional arrangements this is the central bank on its own or in a coordinated action together with the treasury or the ministry of finance) can directly set the nominal exchange rate. This can be put to use because depreciating the currency and pegging it at the depreciated rate serves as a conspicuous commitment to a higher future price level. According to the interest parity condition, the pegging of the currency at a depreciated rate while leaving the rate of interest at a constant (zero) level will set a process in motion wherein the equilibrium real exchange rate is restored by an upward movement of prices. An increase of the present level of the nominal exchange rate therefore signals that long-term inflation

should be expected. This strategy is of course in line with the earlier mentioned 'optimal' response to a liquidity trap. Svensson incorporates this in what he calls a 'foolproof way of escaping from the liquidity trap'. This way requires that the monetary authorities have a firm commitment to a higher future price level, start a concrete action to demonstrate this commitment and define a exit strategy that will be put in place when things get back to normal (Svensson 2003:160-1). Currency depreciation plays a strategic role in this approach because it is the principal factor that induces expectations of a higher price level in the future and enhances the credibility of the announced policies.

The first step is to determine the price level target from which the required depreciation and the corresponding initial exchange rate can be derived. Next the time frame and a target path for return to normality have to be determined. On the basis of this the monetary authorities then announce a crawling peg wherein the initial exchange rate slowly drops in steps equal to the average interest rate differential with the main trading partners. Simultaneously the monetary authorities signal that they will defend the peg by buying and selling any required amount of foreign currency at the announced exchange rate. To make the peg credible the central bank must be prepared to absorb any excess demand for domestic currency. This is easily fulfilled because the central bank can print currency at will. Once the credibility of the peg is established, the initial depreciation and the rational expectation that the relative prices of domestic and foreign goods (the real exchange rate) will return to normality forces the private sector to expect a higher price level in the future.



Although Svenson's 'fool proof way' appears to approximate the required optimal solution to the liquidity trap, there are some caveats. The degree of success of this strategy depends on at least on two factors. The first is the relative size and the degree of openness of the economy that is afflicted by the liquidity trap. The second is the probability that other economies retaliate by a competing depreciation against what is considered a 'beggar they neighbour policy' of the instigator of the strategy.

## Conclusion

It has become the norm for economists to disagree on matters of economic policy. The great variety of opinions about what to do to escape from the liquidity trap clearly illustrates this. There have been attempts to determine the odds of the different approaches by comparing the various proposals to lift the Japanese economy from its liquidity trap (cf. Coenen & Wieland 2003). The results of this effort are, however, ambiguous so the differences of opinion remain. But in a certain sense there is light at the end of the tunnel. It appears that some agreement has been reached regarding the way in which the liquidity trap can be avoided. There is a growing consensus that the setting of an explicit target path for inflation or for the price level is an adequate strategy to pre-empt the dangers of a liquidity trap.

The Japanese example teaches us that the liquidity trap is not some historical myth. It has been a reality in the past and it can become a reality in the future. It is important to try to understand it and to develop the instruments that can be used to avoid it or, in case, to escape from it. It also teaches that there is no straight line to universal truth. You cannot simply discard the theories and policies of the past because, if economic

conditions change, they may again be applicable in the future.

## Selected References

- Barro, R. J., and D.B. Gordon. (1983) 'A Positive Theory of Monetary Policy in a Natural Rate Model', *Journal of Political Economy*, Volume 91, Number 4, pp. 589-610.
- Bernanke, B.S. (2002) *Deflation: Making Sure 'It' doesn't Happen Here*. Boston: Federal Reserve Board.
- Bernanke, B. S. (2003) *Some Thoughts on Monetary Policy in Japan*. Boston: Federal Reserve Board.
- Bernanke, B.S. and M. Gertler. (1995) "Inside the Black Box: The Credit Channel of Monetary Policy Transmission", *Journal of Economic Perspectives*, Volume 9, Number 4, p. 27-48.
- Boianovsky, M. (2004) "The IS-LM Model and the Liquidity Trap Concept: From Hicks to Krugman", *History of Political Economy*, Volume 36, Number 1, pp. 92-126.
- Brunner, K. and A.H. Meltzer. (1988) "Money and Credit in the Monetary Transmission Process", *American Economic Review*, Volume 78, Number 2, pp. 446-451.
- Buiter, W.H. (2001) *The Liquidity Trap in an Open Economy*. CEPR Discussion Papers. No 2923. London: Centre for Policy Research.
- Buiter, W.H. and N. Panigirtzoglou. (1999) *Liquidity Traps: How to Avoid Them and How to Escape Them*. NBER Working Paper 7245. Cambridge, Mass: National Bureau of Economic Research
- Coddington, A. (1997). "Keynesian Economics: The Search for First Principles", in B. Snowdon and H.R. Vane (Editors), *A Macroeconomics*

- Reader*. London and New York: Routledge, pp. 36-54.
- Coenen, G. and V. Wieland. (2003) "The Zero-Interest-Rate Bound and the Role of the Exchange Rate for Monetary Policy in Japan", *Journal of Monetary Economics*, Volume 50, Number 5, pp. 1071-1101.
- Congress, U.S. (1935) *Banking Act of 1935. Hearings before the Committee on Banking and Currency, House of Representatives*. Washington DC: US Congress.
- Eggertsson, G.B. (2003) *How to Fight Deflation in a Liquidity Trap: Committing to Being Irresponsible*. IMF Working Paper No 03/64. Washington DC: International Monetary Fund: IMF.
- Friedman, M. (1968) "The Role of Monetary Policy", *American Economic Review*, Volume 58, Number 1, pp. 1-17.
- Friedman, M. (1977) "Nobel Lecture: Inflation and Unemployment", *Journal of Political Economy*, Volume 85, Number 3, pp. 451-472.
- Friedman, M. and A.J. Schwartz. (1963) *A Monetary History of the United States, 1867-1960*. Princeton: Princeton University Press.
- Goodfriend, M. (2000). "Overcoming the Zero Bound on Interest Rate Policy", *Journal of Money, Credit, and Banking*, Volume 32, Number 4, pp. 1007-1035.
- Hicks, J.R. (1937) "Mr. Keynes and the "Classics"; A Suggested Interpretation", *Econometrica*, Volume 5, Number 2, pp. 147-159.
- Kaldor, N. (1960). "The Radcliffe Report", *Review of Economics and Statistics*, Volume 42, Number 1, pp. 14-19.
- Kalecki, M. (1935) "Stimulating the Business Upswing in Nazi Germany" in J. Osiatynski (Editor), *Collected Works of Michal Kalecki*. (Vol. 6 Studies in Applied Economics, 1927-1941, pp. 196-201). Oxford and New York: Oxford University Press, 1996.
- Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*. London and Basingstoke: MacMillan.
- Kregel, J.A. (2004). "Krugman on the Liquidity Trap: Why Inflation Will Not Bring Recovery in Japan", in R. Arena (Editor), *Money, Credit and the Role of the State. Essays in honour of Augusto Graziani Salvadori Neri*. Aldershot and Burlington: Ashgate, pp. 225-238.
- Krugman, P. R. (1998). It's Baaack: Japan's Slump and the Return of the Liquidity Trap. *Brookings Papers on Economic Activity*, 1998(2), 137-187.
- Krugman, P.R. (2000). "Thinking about the Liquidity Trap", *Journal of the Japanese and International Economies*, Volume 14, Number 4, pp. 221-237.
- Kydland, F.E., and E.C. Prescott. (1977). "Rules Rather Than Discretion: The Inconsistency of Optimal Plans", *Journal of Political Economy*, Volume 85, Number 3, pp. 473-491.
- McCallum, B.T. (2000) "Theoretical Analysis Regarding a Zero Lower Bound on Nominal Interest Rates", *Journal of Money, Credit, and Banking*, Volume 32, Number 4, pp. 870-904.
- McCallum, B.T. (2002) "Inflation Targeting and the Liquidity Trap", in N. Loayza and R. Soto (Editors), *Inflation Targeting: Design, Performance, Challenges*. Series on Central Banking, Analysis and Economic Policies, Volume 5. Santiago: Central Bank of Chile, pp. 395-437.
- McCallum, B.T. (2003) "Japanese Monetary Policy, 1991-2001", *Federal Reserve Bank of Richmond Economic Quarterly*, Volume 89, Number 1, pp 1-31.

- Meltzer, A.H. (1995) "Monetary, Credit and (Other) Transmission Processes: A Monetarist Perspective", *Journal of Economic Perspectives*, Volume 9, Number 4, pp. 49-72.
- Meltzer, A.H. (2001a) "Monetary Transmission at Low Inflation: Some Clues from Japan in the 1990s", *Monetary and Economic Studies*, Volume 19, S1, pp. 13-34.
- Meltzer, A.H. (2001b) "The Transmission Process", in *Deutsche Bundesbank: The Monetary Transmission Process: Recent Developments and Lessons for Europe*. Houndmills, U.K. and New York: Palgrave, pp. 112-30.
- Morgan, B. (1982) *Monetarists and Keynesians*. London and Basingstoke: MacMillan Press.
- Orphanides, A. (2004) "Monetary Policy in Deflation: The Liquidity Trap in History and Practice", *North American Journal of Economics and Finance*, Volume 15, Number 1, pp. 101-124.
- Patinkin, D. (1990) "On Different Interpretations of the General Theory", *Journal of Monetary Economics*, Volume 26, Number 2, pp. 205-243.
- Radcliffe. (1959) *Report of the Committee on the Working of the Monetary System*. (No. Cmd. 827). London: Her Majesty's Stationary Office.
- Samuelson, P.A. (1940) "The Theory of Pump-Priming Re-examined", *American Economic Review*, Volume 30, Number 3, pp. 492-506.
- Sumner, S. (2002) "Some Observations on the Return of the Liquidity Trap", *Cato Journal*, Volume 21, Number 3, pp. 481-490.
- Svensson, L.E.O. (2001) "The Zero Bound in an Open Economy: A Foolproof Way of Escaping from a Liquidity Trap", *Monetary and Economic Studies*, Volume 19, Number 1, pp. 277-312.
- Svensson, L.E.O. (2003). "Escaping from a Liquidity Trap and Deflation: The Foolproof Way and Others", *Journal of Economic Perspectives*, Volume 17, Number 4, pp. 145-166.
- Ullersma, C.A. (2002) "The Zero Lower Bound on Nominal Interest Rates and Monetary Policy Effectiveness: A Survey", *De Economist*, Volume 150, Number 3, pp. 273-297.

Jan Reijnders

School of Economics, Utrecht University

Janskerkhof 12, Utrecht, Netherlands

j.reijnders@uu.nl

## Living Wage

*Oren M. Levin-Waldman*

### **Introduction**

The concept of a living wage dates back to the late nineteenth century when labor leaders in the United States sought to rally organized labor to make wage labor respectable. Then wage labor was viewed no differently than slave labor, and workers saw themselves as being no different than prostitutes. Therefore, the notion that workers could earn wages that would enable them to support their families was one that would also make the enterprise of wage labor respectable. Despite differences in opinion over just what would constitute a living wage, those who espoused it generally made a distinction between it and a mere subsistence wage. A living wage would ideally suffice to support a family. By crusading for a living wage, workers were able to transform the way they viewed themselves and the work they did (Glickman 1997). In time, however, the living wage was to become a foundation for the national minimum wage that was initially adopted in the United States in 1938. While the concept of a living wage was intended to establish a standard for a statutory minimum wage that would serve as a wage floor for all workers, the minimum wage that was ultimately adopted actually fell short of the living wage ideal. In other countries, however, the living wage still aspires to a standard for national minima that will enable low-wage workers to live respectable lives.

The living wage in recent years could be said to reflect the failure of the minimum wage, at least in the United States, to achieve a wage standard, let alone maintain one. And the living wage has certainly been a response to the failure of the minimum

wage to keep pace with inflation. But it has also specifically been a local issue intended to address the needs of those in the low-wage labor market. Living wages, in short, are nothing more than local ordinances that mandate that workers, working for firms with contracts with local governments to perform municipal services, be paid a specified minimum.

Nevertheless, living wages, particularly those in the U.S., need to be understood in their broader context. The living wage has been a measured response to a set of policies that have in part been responsible for these workers being in the position that they would even need such legislation: most notably the privatization and outsourcing of municipal services. Services that were once performed by relatively well-paid and oftentimes unionized employees, are now performed by private sectors contractors that pay their mostly non-unionized workers wages that are at the bottom of the wage distribution.

This essay will look at the living wage in terms of its origins and effects. This essay will also demonstrate the differences between the living wage and the minimum wage and place the living wage in its broader political context. On one level, living wages may represent a quest for economic justice, but on another they are the inevitable byproducts of their respective cities' local political dynamics. In order to understand the concept of the living wage today, one needs to see them within the broader context of local economic development policies, and the politics of local economic development. The living wage movement in the United States is essentially a grassroots backlash against the type of development policies that local governments have pursued. Although pursued out of a perceived need to attract

investment, they were often at the expense of low-wage workers.

### **Living Wage versus Minimum Wage**

The first living wage ordinance was passed by the city of Baltimore in 1994. All companies with contracts to perform municipal services were required to pay a minimum wage of \$6.10 an hour at a time when the federal minimum wage was still only \$4.35 an hour. In the years since, more than 100 municipalities have passed some type of living wage ordinance, and many more are in the process of considering their adoption.

Living wage ordinances differ from both federal and state statutory minimum wages. A minimum wage is generally a wage floor established by a legislative body, whether it be at the state or national level, that mandates that all covered workers be paid that minimum. A living wage only covers those workers employed by firms that have contracts with those municipalities and/or county governments that passed the ordinances. But in some cases living wage ordinances may cover workers employed by firms that have received financial assistance, whether in the form of tax abatements or direct subsidies to invest (ACORN 2001).

In the United States minimum wages are not automatically adjusted according to inflation, rather they only increase following legislative action. Many living wage ordinances are indexed to either the Consumer Price Index (CPI) or increases in state median wages. The typical minimum wage in the U.S. only establishes a wage floor; the typical living wage actually establishes a broader compensation floor. Many living wage ordinances require either the provision of health insurance or paying an even higher wage if no health insurance is provided. Ordinances also often include

provisions for paid vacation and monitoring and enforcement procedures.

Even though living wage ordinances are not as encompassing as minimum wages in their coverage, they seek to offer broader protections. Arguably they seek to offer union style benefits to those who have traditionally been beyond the reach of union organizing, and to the extent that they come about through grassroots organizing campaigns, living wages may be said to offer marginalized workers at the bottom of the wage distribution the type of voice that only unions are able to obtain for their workers.

Despite these differences, however, they both reflect the need for a mechanism to boost wages so that low-wage workers will be able to subsist above the poverty level. Both assume that the market mechanism does not fairly allocate wages, and that when the market fails in this manner wage policy is needed. Moreover, both assume that the benefits of unionism, which have all but eluded low-wage workers, can effectively be extended to the low-wage labor market through legislative action.

The living wage, however, appears to have different meanings in other countries. Campaigns in other countries are actually reminiscent of the historical living wage in the United States, mainly a quest for a national minimum standard. In Canada, for instance, struggles to achieve a living wage have primarily focused on the need to increase the minimum wage, particularly for members of the Homeworkers Association (Eaton and Dagg 2004). In Britain, campaigns for living wages have been about low-wage workers. According to Damian Grinshaw (2004), the British campaign for living wages, or low-pay campaigns as they are called, essentially contain three dimensions: 1) the improvement of the

National Minimum Wage; 2) the elimination of a two-tiered work force; and 3) the prevention of over-reliance on in-work benefits. Because the focus has been on national policy, the achievements have been considerably broader than in the U.S. where the immediate benefits only accrue to those working for firms with local contracts. In Britain, these low pay campaigns have provided new insights for understanding the forces shaping wage structures among low-paid workers. By highlighting the issue of poorly paid workers, they have been successful in pressuring and convincing employing organizations to improve levels of pay.

In Australia, the concept of a living wage is actually an integral element in a national system of wage determination. At the center of Australia's unique wage setting institutions are quasi judicial tribunals that were established around the turn of the century to manage and resolve industrial disputes. The idea of a living wage was a founding concept in the Australian industrial arbitration system. The Australian Council of Trade Unions launched their living wage campaign in 1996 as a claim before the Australian Industrial Relations Commission to vary awards. With the living wage campaign, the ACTU was attempting to mount a counter-offensive against the newly elected government that was committed to breaking union power and deregulating the labor market. The living wage has, in short, been about preventing the collapse of hourly wage rates in an economy plagued by chronic unemployment and underemployment (Buchanan et al 2004).

And in New Zealand, the living wage has been about achieving fair labor standards. In New Zealand's labor market, wage determination occurred principally through centralized bargaining systems until the mid

1980s. "Fair wages" were initially defined with reference to prevailing wage rates, but over time they became conflated with appropriate living standards. For nearly a century, the Industrial Conciliation and Arbitration Act of 1894 had established a basic framework for industrial relations. As part of a 1925 statement of the Arbitration Council, the living wage was considered to be a minimum rate earned by a man that would be sufficient to maintain himself, his wife, and two dependent children. Until 1984, the labor market had been governed by the full-income concept – a social wage – that included all non-wage benefits paid by employers, as well as contributions from government revenue to wage earners or the entire population. But between 1984 and 1990, the Labour party, which was always closely linked to the trade union movement, began a process of dismantling market regulations. In 1991, the National Government enacted the Employment Contracts Act (ECA), which didn't even mention trade unions; rather they were subsumed under bargaining agents, thereby weakening their ability to recruit and represent workers. Given this climate, the living wage in New Zealand has been about improving the national minimum wage, which always fluctuated widely in its ratio to average wages (Heyman 2004).

A key difference, then, between the U.S. and the other countries is that while the living wage aspires to achieve the same ideals, mainly to boost the wages of those at the bottom of the distribution, the strategic goals in the U.S. tend to be narrower. In the U.S. the living wage has been nothing more than an effort to offer assistance to a limited number of low-wage workers at the local level, and very much in the spirit of incrementalism. That is, with a local foundation in place, the precedent exists to

expand further at a later time. Outside the U.S., however, the living wage has been conceived of as a broader effort either to maintain national gains that workers achieved in the past that are perhaps perceived to be under threat, or to obtain a better national wage policy. Arguably, living wage campaigns outside the U.S. have been successful in achieving the ideals found in the rhetoric of justice and fairness employed by many American campaigns to galvanize low-wage workers.

### **Arguments for Living Wages**

The standard political arguments made by living wage campaigns and their leaders are that with the failure of the minimum wage to keep pace with inflation, the outsourcing of municipal services, and a general increase in income inequality, living wages are essential to achieve a measure of justice and fairness. Full-time workers should be paid a wage sufficient to support themselves and their families (Pollin and Luce 1998). Living wages are often touted as anti-poverty measures (Neumark 2004). In this vein, living wages challenge the status quo because they stress the importance of establishing a community standard as to what constitutes a fair wage. They effectively call into question the prevailing notion that wages set by the market place are de facto fair wages (Weldon & Targ 2004).

A central argument for the living wage in the U.S. is that because the economy has failed to generate decent paying jobs sufficient to support families, government must mandate minimum wages to compensate for this structural failure. But because the federal government and most states have failed to do so through their own respective statutory minima, local communities must take it upon themselves to meet the needs of their citizens in the form

of living wages. The federal minimum wage that was passed in 1938 was born at a time when national policy predicated on boosting wages sought to maintain a measure of economic stability. Arguably, the federal minimum wage grew out of an economy undergoing the pains of industrialization and much of the attendant industrial strife. For those who supported the federal minimum, it was only the beginning of what was hoped would be a process intended to achieve a broader living wage (Nordlund 1997; Levin-Waldman 2001).

Higher wages may lead to greater efficiency in that workers become more productive. From an efficiency point of view, employers who pay living wages will realize other savings in the form of reduced turnover, recruitment, and training costs. Insofar as living wages seek to create a wage floor for covered workers some of the analytical criteria applicable to the minimum wage will also apply to the living wage. The standard textbook analysis of the minimum wage has always held that wage floors lead to lower employment because they prevent workers from lowering their wage demands to a level where they might be demanded and ultimately consumed.

Most of the emphasis in the minimum wage literature has been on disemployment effects. But there was always another side to the analytical framework, and one which has not received the same level of attention. That, mainly has been, as pointed out by George Stigler (1946), that if the minimum wage does not lead to less employment, it will lead to greater productivity. Arguments for the living wage in more recent years, however, have been grounded more in efficiency. Indeed, some of the strongest empirical support for the living wage has been in the area of efficiency studies.

Efficiency wage arguments found in mainstream economics hold that workers who receive higher wages have a stronger incentive to hold onto their jobs because the costs associated with job loss are now higher. Employers also benefit because the costs of higher wages are usually offset by savings in monitoring costs (Shapiro and Stiglitz 1990). The growing efficiency wage literature on the living wage tends to follow the more institutionalist approach put forth close to a century ago by Sidney Webb (1912). Webb argued that paying a minimum wage would have the effect of increasing efficiency because workers would become more productive. The employer would have incentive to find ways to increase productivity either by getting workers to produce more or by substituting technology for labor. But the worker would also have incentive to improve his or her skills so that the value of his/her labor would justify the new wage. Moreover, workers would also become more productive because a higher wage would better enable them to maintain themselves physically, which in turn would sustain their morale. The new living wage may similarly be an efficiency wage to the extent that higher labor costs can be offset by savings due to reduced turnover and savings related to recruitment and training (Howes 2002; Fairris 2003; Reich et al 2003).

Finally, the traditional macroeconomic argument for the minimum wage also applies to the living wage. This model suggests that insofar as higher wages offer low-wage workers greater purchasing power, they in turn are able to demand more goods and services. Consequently, businesses are able to produce more, and in turn hire more workers, which may inevitably lead to the natural bidding up of wages across the board. The initial federal minimum wage

was very much informed by this type of macroeconomic argument. Higher wages, it was reasoned, lead to greater stability because they result in less labor strife. Better paid workers, after all, have less reason to engage in labor actions against their employers. Ultimately, the economy as a whole prospers because workers have more money (Prasch 1996). Hence the living wage is further differentiated from the minimum wage by the different foci of the research.

### **Arguments Against the Living Wage**

Arguments against the living wage tend to mirror those against the minimum wage. The living wage effectively creates a wage floor, and according to economic orthodoxy the costs to society of imposing wage floors are greater than any benefits. According to the theory of Perfectly Competitive Markets, market clearing wages are achieved when the demand for labor is exactly equal to the supply of labor. In such a market, there is no such thing as unemployment because wages either rise or fall until the demand for labor is exactly equal to the supply of labor. At the wage at which demand equals supply, all those willing and able to work at that wage will be employed. If more people are willing to work, the wage will fall further, thereby inducing firms to hire more workers, with the result being that the supply of labor once again equals the demand. Conversely, when firms are unable to hire as many workers as they would like, the wage rises to induce additional people to enter into the workforce until supply and demand are once again equal.

A wage floor, such as a mandated minimum wage, or a living wage, prevents the cost of labor from dropping below that minimum. When the wage floor is higher than the equilibrium wage, fewer workers will be hired than are willing to work,



thereby resulting in unemployment. In a competitive market, each worker receives the value of his or her marginal revenue product, which is the amount of increase in the output that results from an increase in say a unit of labor. If adding an additional worker results in a rise in total revenues, the firm's output will rise as a result. Firms typically use the marginal revenue product of labor as a criterion for determining how many more workers to hire because they are able to calculate how much more output can be expected based on how many units they add. Therefore, a wage floor will result in the layoff of those workers whose value is less than the minimum (Stigler 1946; Ehrenberg & Smith 1997). Employers either lay workers off to compensate for higher labor costs, don't create new jobs in the future, or substitute either technological improvements or workers with greater skills that can result in greater efficiency. The possibility also exists that employers will cut back on training and quality of life, i.e. non-wage types of compensation, in the workplace (Prasch & Sheth 1999). Therefore, a wage floor only ends up hurting low-wage workers—precisely those whom it was intended to help.

Arguments against living wage ordinances per se tend to focus on the impact they have on their respective cities. The Employment Policies Institute (2000), for instance, has argued that living wage activists do nothing more than offer emotional rhetoric against economic reality. A living wage is thus nothing less than an organized effort to force some employers to pay wages based on a definition of “need” rather than “skills.” David Neumark and Scott Adams (2000) argue that living wage ordinances may not only be restricted in coverage to only those workers who are performing municipal services in fulfillment

of a contract, but may extend to other employees in these same firms working on contracts not covered by living wage ordinances. Firms forced to pay higher wages to employees on a city contract may feel pressured to pay higher wages to their other employees not working on city contracts. That is, living wages may have wage contour effects (Neumark et al 2004). But unlike the contractor with the public sector that s/he may be able to pass on the costs to, s/he cannot do that with firms in the private sector.

Because the living wage applies only to private contractors, it will adversely affect the competitive bidding process. Even if no one is laid off, contractors will still need to pass on the costs thereby raising their costs and the costs of municipal budgets. Or worse, they simply opt out of the bidding process, thereby leaving fewer in the pool to bid, which will also have the consequence of raising prices. Living Wages, it is alleged, can harm a city's economic base by sending investors signals that it is an unfavorable business climate. Critics in Baltimore, for instance, claimed that higher wages on city contracts would only worsen the city's already precarious fiscal position with negative consequences for tax rates and the provision of municipal services. Investors would simply opt to invest in the surrounding suburban communities where such ordinances did not exist (Neidt et al 1998). As such, a living wage is no more than a “leaving wage” because it may force employers, employees, customers and taxes to desert a given city (Walters 1994). Moreover, a living wage ordinance, especially were it to create a precedent for a more general citywide minimum wage, would only hamper the city's ability to attract private investment.

The living wage, then, is likely to have what Neumark and Adams (2000) refer to as both “first round” and “second round” effects. In the first round, firms “select out” of the city, and those most likely to do so are those firms that have the highest share of low-skilled labor. Hence low-skilled workers who lose their jobs because of the process of selecting out are precisely those that the ordinances were intended to help in the first place. The second round, then, results from the first. As some firms terminate contracts with the city, there are then fewer firms left to bid on city contracts, which only results in less competitive bidding and therefore higher prices for cities (Neumark & Adams 2000). A living wage, it is argued, hinders efforts to attract investment because it effectively sends the wrong signals to would-be investors, suggesting that the city that passes such an ordinance is an unfavorable business climate (Levin-Waldman 2005).

### **Effects of Living Wages**

Unlike the minimum wage where there is considerable literature on the labor demand effects, at least on the teen labor market (Kosters & Welch 1972; Welch 1974, 1978; MWSC 1981; Meyer & Wise 1983; Neumark & Wascher 1992; Kosters 1996), there is little research thus far on the living wage. The key issues are 1) effects on poverty and employment; 2) costs to municipal budgets; 3) costs to employers as measured against benefits to workers; 4) whether contractors seek to drop municipal contracts because of rising costs; and 5) whether there are efficiency wage benefits to employers and cities. Using data from the Current Population Survey (CPS) for the years 1996-2000, David Neumark (2004) found that poverty in those cities that passed living wage ordinances was less than in

those cities that did not pass such ordinances. At the same time, however, there were sizeable wage gains above the 10<sup>th</sup> percentile level, and there were actually positive employment effects between the 50<sup>th</sup> and 75<sup>th</sup> percentiles (pp. 40-86). The problem with these findings, however, is that the best that he can offer is a positive correlation. He cannot establish that poverty was in fact lower because of the living wage, which is to say that he cannot establish a causal relationship. Moreover, because his econometric model relies on a truncated sample of workers which excludes higher wage workers, there is the problem of sample selection bias. Moreover, the CPS is individual level data and therefore it cannot capture the actual experiences of those cities that have adopted these ordinances (Brenner et al 2002).

What about the budget impacts? Much of the literature to date is prospective insofar as it tends to estimate the effects of ordinances on cities in terms of their bidding and service contracts. This literature generally estimates the costs to be low, and that employment consequences will also be relatively low (Pollin & Luce 1998; Nissen & Cattani 1998; Benner & Rosner 1998; Zabin et al 1999; Reynolds et al 1999). Robert Pollin (2003), maintains that it does not necessarily follow that higher wage costs will necessarily be passed onto municipal budgets dollar-for-dollar in the form of more expensive service contracts. Firms competing for municipal contracts are more likely to absorb the rising costs because they do not want to price themselves out of the bidding process. Christopher Neidt et al (1998) found the budgetary impact on the city of Baltimore to be generally insignificant. In the area of grass cutting, for instance, some medium sized contractors withdrew, thereby prompting the subdivision

of contracts in favor of small-scale management. In bus contracting, however, there was a slight shift toward the larger, national companies and away from the smaller local companies. But they also found that when they adjusted for inflation, the cost of a set of comparable contracts from before and after the ordinance took effect actually declined slightly.

In their study of the impact of the ordinance at San Francisco International Airport, Michael Reich et al (2003) found that ordinances, which were part of larger Quality Standards Program (QSP) lead to less turnover. Turnover fell by an average of 34 percent among all surveyed firms and 60 percent among those firms where average wages increased by at least 10 percent. In fact, following a fifteen month period after the implementation of QSP, turnover fell by almost 80 percent from 94.7 percent to 18.7 percent. Because employers were paying an average of \$4,275 every time a worker had to be replaced, employers ended up saving around \$6.6 million each year in turnover costs, despite an increase in overall labor costs. To the extent that employers experienced reduced turnover costs, they experienced productivity gains. Many reported that the quality of work increased, and many workers themselves indicated that they were more inclined to put more effort into their work. Moreover, there appeared to be no significant reductions in employment; rather employment at the airport actually increased by around 15.6 percent during the period in which QSP was implemented.

### **The Living Wage in Broader Perspective**

Despite the growing literature on living wage effects from an economics viewpoint, and even the perspective of union organizing (Reynolds 1999, 2001; Ciscel 2000), the bulk of the literature on the living wage is

simply too narrow. Because living wages in the U.S assume the form of municipal ordinances, they are specifically local issues and therefore need to be understood within the context of urban politics and the role that cities play in the larger national framework (Levin-Waldman 2005).

Politics at the local level are often characterized by a competitive struggle to attract investment and hence revolve around issues of development and redevelopment (Peterson 1981; Eisinger 1988). To attract investment, local politicians often enter into partnerships and form coalitions with business interests, and these coalitions are often what define the nature of a local governing regime (Stone 1989; Stoker 1995). These regimes often revolve around growth, which inevitably flow from cities' economic dependence on private investment, as they are forced to compete with one another to attract capital investment (Molotch 1976; Kantor 1995). Therefore, local governments are loathe to pursue policies of benefit to low-income residents because they are essentially redistributive, and hence send the wrong signals to potential investors (Peterson 1981). On the contrary, local officials often feel compelled to create a favorable business climate, which often entails reducing municipal expenditures. The contracting out of municipal services has been but one way of sending positive signals to investors.

Andy Merrifield (2002) argues that the living wage is best understood within the broader political context of power, conflict, and protest that make up the dialectics of urban life. When cities pursue development, because they must out of necessity, there is no escaping that it will set off a movement of conflict and protest, and the living wage is but one manifestation of that (p. 78). Indeed, the Baltimore living wage ordinance grew

out of a campaign led by a coalition of community groups and civic organizations to assist low-wage workers, many of whom may have formerly been performing municipal services on the city payroll. The pattern was all too familiar: work that had been done in the past by unionized municipal workers was now being outsourced to private contractors, who in many cases were not paying their workers much more than the federal minimum wage.

The living wage is thus the inevitable byproduct of urban politics and their evolution in the United States. To the extent that there are dynamics to urban politics that have been set into motion because of external forces, i.e. the unique position that cities find themselves in, the living wage movement is only a continuation of those dynamics. Are there some cities, then, that are more likely to pass ordinances than others? Can we identify a set of conditions that need to be in place in order for a living wage campaign to be successful?

### **City Ordinances**

Elsewhere (Levin-Waldman 2004a, 2004b), I have examined some of those features on the basis of data from the Current Population Survey (CPS) by comparing cities that passed ordinances to those that did not. I specifically looked at these two groups of cities at two points in time: 1993, the year before the first living wage ordinance was enacted, and 2002. Cities that passed ordinances had higher levels of income inequality in 1993 prior to passage of the first ordinance. Although the passage of ordinances did not reverse the trend, the percentage growth did appear to be less in those cities that passed them over those that did not. Whereas family income inequality rose by 18.42 percent between 1993-2002 in cities that did not pass ordinances and by

18.75 percent in the rest of the country, it only rose by 10.53 percent in those cities that did pass ordinances. Although it cannot be established that the lower percentage growth in inequality was due to the ordinances per se, it can be speculated that the lower percentage growth was perhaps due to the fact that in those cities the percentage increase in family income of those at the bottom of the distribution was higher than those at the top.

Cities that passed ordinances had higher immigrant populations and lower levels of educational attainment. Differences between living wage and non-living wage cities were especially great among Mexican Americans, Mexicans, Central or South Americans, and Cubans. Therefore, to the extent that there are more immigrants in living wage cities, and they came with little skills, the greater inequality in those cities might be accounted for by the fact that an influx of individuals with lower skills will result in greater inequality because their increasing numbers effectively suppresses wages at the bottom of the distribution (Nord 1980).

Living wage cities also had more people employed in both the Durable and Non-Durable goods industries, as well as services. Therefore, to the extent that there is a relationship between low educational attainment and low wage levels, or the greater probability of being employed in low wage occupations, we might infer that the wages of those at the bottom are lower in those cities that passed ordinances because educational levels are lower, thereby implying a greater need for labor market institutions—institutions that will boost wages. Consequently, these cities also effectively have sizeable labor markets at the low end of the income distribution that are ripe for organizing. Therefore, it might be plausible to infer that in order for a city to be

predisposed towards passing such ordinances, there must at a minimum be a population that stands to benefit. An immigrant population stands to benefit because it lacks the requisite education and skills to command higher wages. An immigrant population also stands to benefit because it is perhaps being exploited. That is, the structure of the labor market, and the demographics of that labor market, may predispose it more to the adoption of such ordinances. Consequently, political campaigns for living wages that occur at the local level must be seen as a response to the changing economic circumstances, and are intimately connected to the demographics of their respective communities.

## **Conclusion**

In conclusion, then, we might say that living wages occur within a context of changing economies—from an industrial, manufacturing to post-industrial service sector—rising income inequality, the outsourcing of municipal services, and an urban political culture predicated on growth. The politics of development and redevelopment, from which the living wage arises, is certainly driven by larger economic transformations. Therefore to the extent that the minimum wage grew out of the need to respond to all the problems attendant to industrialism, the living wage might be said to represent the next stage in that general evolution, as it arises from the post-industrial transformation. Moreover, cities with certain demographics, particularly large immigrant populations and low educational attainment are thus more likely to have sizeable labor markets at the low-end of the income distribution. As such, these cities may only offer greater opportunity for living wage campaigns to organize these workers, who, because they are at the bottom of the

wage scale, also have greater incentive to join in living wage campaigns because they are in a position to benefit. In other words, growing income inequality in urban labor markets, particularly with specific demographic profiles, may, because of the nature of those profiles, effectively create incentives for collective action to launch living wage campaigns.

These lessons from the U.S. would also appear to have implications for developing countries that have not as of yet suffered the same pangs of economic transformation. The quest for a living wage is clearly a response to economic transformations. As developing countries continue to develop their economies, there will most likely be ample opportunities for living wage campaigns to emerge as the logical antidote to the process. That the living wage is in response to the failure of the minimum wage in the U.S., and that it reflects a desire to increase the national minimum wage elsewhere, it perhaps serves as an important benchmark for what national minima should be. In an era of diminishing unionization, the living wage represents an attempt to include more workers under the umbrella of protections once only enjoyed by unionized members. It does give voice to low-wage workers. But herein lies its greatest challenge --- to extend coverage. In countries where living wages do not exist, it could serve the important function of galvanizing workers and sensitizing populations to issues of pay equity.

In the U.S. where the living wage movement has emerged as a series of campaigns that have successfully passed a variety of ordinances, the challenge is to make them the basis for national policy. What campaigns in the U.S. shares in common with other campaigns around the world is the desire to either achieve a social

wage or reinstate one that has been eliminated. Living wage campaigns, however, need to be about more than the needs of low-wage workers; rather they have to speak the language of the middle class. The living wage has to be conceived of as broader working class issue, and not something that only affects low-wage workers. The living wage needs to set the standard for labor market institutions that can compensate for the decline of labor unions in industrialized countries and the absence of those organizations in developing ones.

Short of this, however, the living wage movement, at least as it has taken place in the U.S. and some other countries, perhaps sets the standard for labor organizing. It means that organizing workers, particularly the marginalized at the bottom of the distribution, is alive and well. And the model of organizing—the grassroots campaign for support—is one that can be duplicated in other countries where general labor protections, let alone living wage laws, do not exist.

### Selected References

- ACORN. (2001) “Living Wage Ordinances ‘Stats at a Glance’” (March 21). downloaded from [www.livingwagecampaign.org](http://www.livingwagecampaign.org).
- Benner, Chris and Rachel Rosner. (1998) *Living Wage: An Opportunity for San Jose*. Working Partnership USA.
- Brenner, Mark D., Jeannette Wicks-Lim and Robert Pollin. (2002) *Measuring the Impact of Living Wage Laws: A Critical Appraisal of David Neumark’s How Living Wage Laws Affect Low-Wage Workers and Low-Income Families*. Amherst, Mass: Political Economy Research Institute. Working Paper Series Number 43.
- Buchanan, John, Ian Watson and Gabrielle Meagher. (2004) “The Living Wage in Australia: History, Recent Developments, and Current Challenges”, In Deborah M. Figart (Editors), *Living Wage Movements: Global Perspectives*. London and New York: Routledge.
- Ciscel, David H. (2000) “The Living Wage Movement: Building a Political Link from Market Wages to Social Institutions”, *Journal of Economic Issues*, 34,2 (June):527-535.
- Eaton, Jonathan and Alexandra Dagg. (2004) “Organizing Homeworkers in Toronto’s Garment Industry”, In Deborah M. Figart (Editors), *Living Wage Movements: Global Perspectives*. London and New York: Routledge.
- Ehrenberg, Ronald G and Robert S. Smith. (1997) *Modern Labor Economics: Theory and Public Policy*. Sixth edition. Reading, MA: Addison-Wesley.
- Eisinger, Peter K. (1988) *The Rise of the Entrepreneurial State: State and Local Development Policy in the United States*. Madison: University of Wisconsin Press.
- Employment Policies Institute. (2000) *The Living Wage Policy: The Basics*. [www.LivingWage.org](http://www.LivingWage.org)
- Fairris, David. (2003) “The Impact of Living Wages on Employers: A Control Group Analysis of the Los Angeles Ordinance”, Living Wage Research Conference. University of California, Riverside, CA (April 11-12).
- Glickman, Lawrence B. (1997) *A Living Wage: American Workers and the Making of Consumer Society*. Ithaca and London: Cornell University Press.
- Grimshaw, Damian P. (2004) “British Living Wage and Low Pay Campaigns”, In Deborah M. Figart (Editors), *Living Wage Movements: Global Perspectives*. London and New York: Routledge.

- Howes, Candace. (2002) "The Impact of a Wage Increase on the Workforce Stability of IHSS Home Care Workers in San Francisco County", *Working Paper*. University of California Institute for Labor and the Economy and the University of California, Berkeley, Center for Labor Education and Research. November.
- Hyman, Prue. (2004) "The Fight for Living Standards in New Zealand", In Deborah M. Figart (Editors), *Living Wage Movements: Global Perspectives*. London and New York: Routledge.
- Kantor, Paul. (1995) *The Dependent City Revisited: The Political Economy of Urban Development and Social Policy*. Boulder, CO: Westview Press.
- Kosters, Marvin H. (1996) (Editor) *The Effects of the Minimum Wage on Employment*. Washington DC: AEI Press.
- Kosters, Marvin H. and Finis Welch. (1972) "The Effects of Minimum Wages on the Distribution of Change in Aggregate Employment", *American Economic Review*, Volume 62, Number, pp. 323-332.
- Levin-Waldman, Oren M. (2005) *The Political Economy of the Living Wage: A Study of Four Cities*. Armonk: M.E. Sharpe.
- Levin-Waldman, Oren M. (2004a) "Exploring the Demographic Factors Affecting Passage of Living Wage Ordinances", *Working Paper # 88*. University of Massachusetts, Amherst: Political Economy Research Institute.
- Levin-Waldman, Oren. (2004b) "Cities that Pass Living Wage Ordinances", *Challenge: The Magazine of Economic Affairs*, Volume 47, Number 5, pp. 56-68.
- Levin-Waldman, Oren. (2001) *The Case of the Minimum Wage: Competing Policy Models*. Albany: State University of New York Press.
- Merrifield, Andrew. (2002) *Dialectical Urbanism: Social Struggles in the Capitalist City*. New York: Monthly Review Press.
- Meyer Robert H. and David A. Wise. (1983) "The Effect of the Minimum Wage on the Employment Earnings of Youth", *Journal of Labor Economics*, Volume 1, Number 1, pp. 66-100.
- Minimum Wage Study Commission. (1981) *Report of the Minimum Wage Study Commission Vol 1: Commission Findings and Recommendations*. Washington DC: Government Printing Office.
- Molotch, Harvey. (1976) "The City as Growth Machine", *American Journal of Sociology*, Volume 82, Number 2, pp. 309-355.
- Neidt, Christopher, Greg Reuters, Dana Wise and Erica Schoenberger. (1998) "The Effects of the Living Wage in Baltimore", *Working Paper 119*. Economic Policy Institute (February).
- Neumark, David, Mark Schweitzer and William Wascher. (2004) "Minimum Wage Effects throughout the Wage Distribution", *Journal of Human Resources*, Volume 39, Number 2, pp. 425-450.
- Neumark, David. (2004) "Living Wages: Protection for a Protection from Low-Wage Workers?", *Industrial and Labor Relations Review*, Volume 58, Number 1, pp. 27-51.
- Neuman, David. (2002) *How Living Wage Laws Affect Low-Wage Workers and Low-Income Families*. San Francisco: Public Policy Institute.
- Neuman, David and Scott Adams. (2000) "Do Living Wage Ordinances Reduce Urban Poverty?" *Working Paper 7606*. Cambridge, Mass: National Bureau of Economic Research (March).

- Neuman, David and William Wascher. (1992) "Employment Effects of Minimum and Subminimum Wages: Panel Data on State Minimum Wage Laws", *Industrial and Labor Relations Review*, Volume 46, Number 1, pp. 55-81
- Nissen, Bruce and Peter Cattani. (1998) *The Impact of a Living Wage Ordinance in Miami-Dade County*. Miami: Florida International University, Center for Labor Research and Studies. (October 23).
- Nord, Stephen. (1980) "An Empirical Analysis of Income Inequality and City Size", *Southern Economic Journal*, Volume 46, Number 3, pp. 863-872.
- Nordlund, Willis J. (1997) *The Quest for a Living Wage: The History of the Federal Minimum Wage Program*. Westport, CT: Greenwood Press.
- Peterson, Paul E. (1981) *City Limits*. Chicago: University of Chicago Press.
- Pollin, Robert. (2003) "Evaluating Living Wage Laws in the United States: Good Intentions and Economic Reality in Conflict", *Working Paper Series #61*. University of Massachusetts, Amherst: Political Economy Research Institute.
- Pollin, Robert and Stephanie Luce. (1998) *The Living Wage: Building a Fair Economy*. New York: The New Press.
- Prasch, Robert E. and Falguni A. Sheth. (1999) "The Economics and Ethics of Minimum Wage Legislation", *Review of Social Economy*, Volume 57, Number 4, pp. 466-487.
- Prasch, Robert E. (1996) "In Defense of the Minimum Wage", *Journal of Economic Issues*, Volume 30, Number 2, pp. 391-397.
- Reich, Michael, Peter Hall and Ken Jacobs. (2003) *Living Wages and Economic Performance: The San Francisco Airport Model*. Institute of Industrial Relations. University of California, Berkeley, CA. March.
- Reynolds, David B. (2001) "Living Wage Campaigns as Social Movements: Experience from Nine Cities", *Labor Studies Journal*, Volume 26, Number 2, pp. 31-65.
- Reynolds, David B. (1999) "Coalition Politics: Insurgent Union Political Action Builds Ties between Labor and the Community", *Labor Studies Journal*, Volume 24, Number 3, pp. 54-75.
- Reynolds, David B. with Rachel Pearson and Jean Vortkamp. (1999) *The Impact of the Detroit Living Wage Ordinance*. Center for Urban Studies and Labor Studies Center, College of Urban, Labor and Metropolitan Affairs, Wayne State University (September 21).
- Shapiro, Carl and Joseph E. Stiglitz. (1984) "Equilibrium Unemployment as a Worker Discipline Device", *American Economic Review*, Volume 74, Number 3, pp. 433-444.
- Stigler, George J. (1946) "The Economics of Minimum Wage Legislation", *American Economic Review*, Volume 36, June, pp. 358-365.
- Stoker, Gerry. (1995) "Regime Theory and Urban Politics", in David Judge, Gerry Stoker and Harold Wolman (Editors), *Theories of Urban Politics*. Thousand Oaks: CA: Sage Publications.
- Stone, Clarence N. (1989) *Regime Politics: Governing Atlanta, 1946-1988*. Lawrence: University Press of Kansas.
- Walters, Stephen J.K. (1994) "Is BUILD Trying to Tear Down?", *Baltimore Sun* (June 22).
- Webb, Sidney. (1912) "The Economic Theory of a Legal Minimum Wage", *Journal of Political Economy*, Volume 20, Number 10, pp. 973-998.



Welch, Finis. (1974) "Minimum Wage Legislation in the United State", *Economic Inquiry*, Volume 12, Number 3, pp. 285-318.

Weldon, S. Laurel and Harry Targ. (2004) "From Living Wages to Family Wages?", *New Political Science*, Volume 26, Number 1, pp. 71-98.

Zabin, Carol, Michel Reich and Peter Hall. (1999) *Living Wages at the Port of Oakland. Center for Labor Research and Education, Center on Pay and Inequality. Institute of Industrial Relations, University of California-Berkeley* (December).

*Oren M. Levin-Waldman*  
*Metropolitan College of New York*  
*New York City*  
*OLEvin-Waldman@mcny.edu*

## Long Waves of Growth and Development

*Vincent Barnett*

### **Introduction**

Long waves of economic development are sometimes referred to as long cycles, major cycles or Kondratiev waves (K waves). They are hypothesized to be approximately fifty to sixty-years in duration, and are constituted from alternating periods of expansion and stagnation across the world capitalist economy. It is suggested that these major cycles of growth and recession have been synchronous across many countries, and have been observable in a number of economic variables (such as price levels and interest rates) since the end of the eighteenth century (Goldstein 1988:21).

Long wave theorists assert that their motion is caused either by the amortisation of basic capital goods; the diffusion of technological styles; the links between production, military expenditure and prices; the structural progression of capitalist crises; phases of institutional evolution; or some other cluster of factors depending on the particular version of the concept being considered. Long cycles are sometimes viewed as being analogous to the eleven-year business cycle accepted as existing by most economists.

However, long waves are highly controversial, having been characterized by Paul Samuelson as ‘science fiction’, and many mainstream economists deny emphatically that they actually exist. Consequently, a number of competing theories have been developed to explain the empirical observations that have led some to posit the existence of long waves as real economic cycles.

According to Wesley Mitchell’s classic definition, a business cycle was a type of

fluctuation found in the aggregate economic activity of nations, that consisted of expansions and contractions in business performance in a sequence that was recurrent but not strictly periodic (Burns and Mitchell 1946:3). Using this general definition, long waves could possibly qualify as being Mitchell-type cycles, depending on the precise definition of ‘recurrent’ that was being applied.

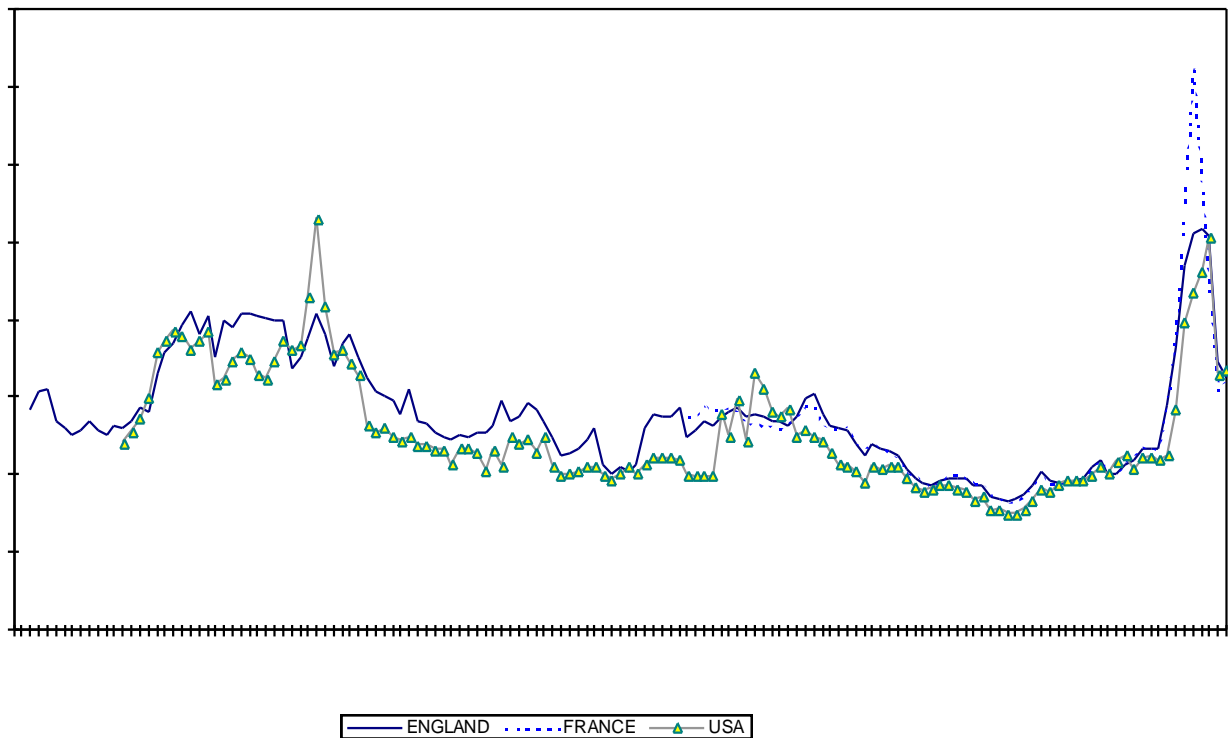
Two basic forms of the long wave hypothesis can usefully be articulated. A stronger version claims the existence of fifty-year long cycles in the capitalist economy that progress with approximate regularity. A weaker form suggests that long waves are more appropriately conceived as temporally unique secular swings of business activity that are generated by a structured interaction of the various economic and technological factors involved. Consequently, the term “long cycles” will generally be used to connote the stronger form of the hypothesis, and the term “long waves” to connote the weaker form. Whether either form falls neatly within Mitchell’s definition cannot be definitively answered here.

### **The Founders**

Various precursors of long wave analysis have been documented, such as Hyde Clark (1847), W.S. Jevons (1884) and J. Van Gelderen (1913), but the true ‘father’ of long cycles was Nikolai Kondratiev, a Russian agricultural economist who became a tragic victim of Joseph Stalin’s political purges. In the early 1920s Kondratiev first searched for long cycles in price indices for various countries such as the UK, the USA and France within the context of the embryonic econometric practices of the Conjunction Institute in Moscow.

According to Kondratiev, long cycles were composed of two segments of similar duration, a rising wave and a declining wave, analogous to the upswing and downturn phases of the business cycle (Barnett 1998b:107). In his first article devoted to this topic published in 1925, Kondratiev outlined the long cycle chronology as follows. The rising wave of

the first long cycle lasted from 1789 to 1814, and the declining wave of this cycle from 1814 to 1849. The rising wave of the second long cycle lasted from 1849 to 1873, the declining wave from 1873 to 1896. The rising wave of the third long cycle began in 1896 and ended in 1920, after which the declining wave began.



*Source:* Kondratiev (1925)

This particular version of cycle periodicity was accepted by some long wave theorists but has been modified and even fundamentally disputed by others. Kondratiev himself developed the idea of temporal bands or ‘probable limits’ for dating the long cycle turning points, as data sets from different countries did not always indicate an exact match for either peak or trough.

From its inception Kondratiev’s analysis was international in scope, and it highlighted fluctuations in key sectors of the national

economies that were involved, sectors such as iron, coal, agricultural wages and foreign trade. The most frequently reproduced graph depicting the long cycle showed indices of commodity prices from the 1780s to the mid-1920s, in which three long cycle ‘peaks’ could clearly be identified. The fact that price indices were often employed to initially demonstrate the long cycle was however controversial.

Other features identified in early studies were that the standard business cycle upturns were longer within the rising wave of the

long cycle and shorter in the declining wave; that gold output grew and more countries were drawn into the orbit of the world capitalist economy in the early period of the rising wave; and that agriculture suffered especially during the declining wave of the long cycle (Kondratieff 1928; Kuznets 1967:264).

Kondratiev's embryonic theory of long cycles focused on basic capital goods, and suggested that large-scale capital investments occurred in distinct wave patterns as such investment goods wore out and/or became obsolete. The type of investment projects being envisaged were railways, large building schemes such as dams, and the technological renovation of factories. Kondratiev also borrowed his tutor, M.I. Tugan-Baranovsky's, notion of free loanable capital being accumulated in the downswing phase of the cycle and then being exhausted in the upswing, and applied it in modified form to long waves (Barnett 2001:454).

In a later article, published in 1928 Kondratiev modified his account, conceiving of the long cycle as a cycle of relative prices between industrial and agricultural goods, caused by the differing dynamics of labour productivity growth in these two sectors of the economy. However Kondratiev readily acknowledged that his own explanation of long cycles was far from complete, and he was prevented from further developing his analysis after 1930 due to the stringent conditions of his imprisonment. In 1938 he was executed. Subsequently various theorists have provided more substantial accounts as follows.

### **Later Expansions of Long Cycle Analysis**

A particularly interesting early attempt to follow up on the work of the long cycle pioneers was undertaken by one of

Kondratiev's colleagues in the Conjunction Institute, T.I. Rainov. Rainov conducted a detailed analysis of variations in the level of scientific discoveries (in particular new discoveries in physics) during the eighteenth and nineteenth centuries in various countries such as England, France and Germany. Rainov claimed that the number of such discoveries exhibited a distinct tendency which, after applying certain statistical procedures, revealed a long temporal fluctuation of several decades.

He concluded that long swings of creative productivity lasting several decades did actually exist (Rainov 1929:300). Various criticisms of Rainov's work were made from within the Conjunction Institute, such as the fact that the timing of new discoveries in basic research was not necessarily linked to their applied use in new technologies, and that expressing all scientific discoveries in the same unit of account (without concern for their relative importance) was problematic.

The most significant early advocate of long cycles was Joseph Schumpeter. In the late 1930s Schumpeter identified a model of capitalist economic evolution involving three cycles—Kondratiev or long cycles lasting around forty five to sixty years; Juglar or intermediate cycles lasting seven to eleven years; and Kitchen or short cycles lasting three to five years. Each Kondratiev contained an integral number of Juglars and each Juglar an integral number of Kitchens (Schumpeter 1939: vol.1, 212). The absence of Kuznets cycles (of between 15 and 25 years duration) from this schema was noticeable.

In Schumpeter's account long cycle upswings were generated by basic technological inventions (such as the steam engine) occurring alongside the 'swarming' of smaller, secondary innovations. These

smaller inventions were initially bunched in the downswing phase, but then they were absorbed into the economy at an increasing rate, as entrepreneurs became more adventurous in their assessment of the potential profits from the fresh business application of these inventions (Tylecote 1991:15).

Schumpeter employed a definition of innovation as the creation of a new production function, or the establishment of a new way in which the quantity of product varied, if the quantity of factors employed in production had altered. In Schumpeter's view innovation was not the same as scientific invention. Innovation could involve various elements, such as technological change; the opening up of new markets or new sources of supply; an improved handling of materials; new business organisations; and so on. Capitalist economic evolution was thus the process brought about by innovation in a broad sense, and the observed multiplicity of business cycles was the corollary of the interactive absorption of innovation into the economic system over time (Schumpeter 1939: vol.1, 86-87).

From a quite different perspective, W.W. Rostow developed his own interpretation of the long cycle idea by positing that they were a measure of secular changes in the relative prices of basic commodities as against manufactures. Long cycles in the price level were thus rooted in periods of the relative scarcity of primary goods (foodstuffs and agricultural raw materials) as against industrial products (Rostow 1990:470-72). One factor linked to this observation was that technical progress in manufactures tended to be more rapid than it was in agriculture, leading to additional forces causing shifts in price relations between the two sectors over time. The

Rostow approach to long cycles had important policy consequences for those countries that specialised in the production of agricultural as against industrial goods. Rostow also attempted to analyse the forces acting on the trade relations between advanced capitalist countries and less developed regions, arguing that the controversial Prebisch-Singer hypothesis of a declining agricultural-manufacture terms of trade was not generally correct.

A quasi-monetarist explanation of long cycles can be developed from the approach of economists like Milton Friedman. Friedman viewed the long cycle phenomenon as being generated by periods of stable rates of growth, that were then interrupted by episodes of significant monetary instability, from which the financial system eventually rebounded (Friedman & Schwartz 1963:187). Others have proposed the notion of long uptrend and long downtrend sequences in US wholesale prices which were attributable to money and credit creation and related gold discoveries and wars (Zarnowitz 1992: 237-38). In this type of approach there can be no temporal regularity to long waves, as the timing of turning points was determined exogenously.

This approach has as its origin Walter Layton's attempted explanation of long cycles based on the increased availability of precious metals. Layton had associated long-term movements in the wholesale price index with changes in the level of gold production and additional discoveries of gold in new world countries like the USA and Australia (*a la* Kondratieff). Thus from within the perspective of the monetarist interpretation of long waves, the consequence of monetary policy and of monetary policy regimes was seen as crucial.

Finally, a lineage of Marxian analysis of long cycles can be traced from Leon Trotsky onwards. Trotsky had articulated the notion of a long-period curve of capitalist development in which superstructural phenomena such as wars and also imperialist expansion generated ascending or declining epochs of industrial progress. More recent efforts in the Marxian vein have focused on changes to the rate of profit and the rate of surplus value.

These approaches posited that Marx's version of the law of the tendency of the rate of profit to fall (caused ultimately by a rising organic composition of capital) acted to precipitate the downswing phase of the long cycle, but that a clustering of political defeats of the working class subsequently acted to offset this tendency. These political defeats produced temporary upturns in the rate of profit and consequently underpinned the ascending phase of the long cycle (Mandel 1976). Within the general Marxian framework long cycles were frequently seen as controversial, due to a problematic relation with more orthodox themes (such as those of the capitalist breakdown controversy) that dominated Marxist economics for many decades.

### **Theories of Long Waves**

Significant augmentations of the theory of long cycles have been accomplished more recently by various authors such as Ernest Mandel, Chris Freeman, Brian Berry, David Gordon and Andrew Tylecote. Such further developments have not always been compatible with each other (or with the ideas of the founders) in terms of methodology and/or results. Still other commentators such as Solomos Solomou have provided more sophisticated criticisms of the long cycle idea

Adapting ideas from 'conventional' business cycle theory to obtain a taxonomy of approaches, explanations of long waves can generally be described as exogenous, endogenous or a combination of both. Exogenous explanations cite new discoveries of gold or new technological inventions as 'shocks' outside the system that cause the long-run equilibrium path of an economy to be disturbed. Endogenous explanations cite structural factors such as the accumulation and exhaustion of funds for investment in basic capital goods, or the manner in which new technologies become dispersed and then replaced among individual enterprises, as conditions that generate the observed long waves of economic growth. More sophisticated explanations might include both of these elements in some mutually interactive form.

Some authors argued for long waves to be understood in terms of a sequence of eras of economic progress. Each era was marked by the development of a cluster of technologies that were supported by institutional structures and that acted to drive economic growth. Examples of such clusters of technologies were the water-powered mechanisation of industry, the electrification of the economy, and the computerisation of networked systems; the development of each cluster was said to constitute an industrial and/or a technological revolution (Freeman and Louçã 2001:140-42). Existing firms often found it difficult to overcome the inertia of their previous path dependent trajectories between such eras of economic development. Hence new firms frequently emerged from scratch to meet the challenge of harnessing the new technology within a business environment, despite some generic entry barriers to innovation such as the need for significant R&D capacity.

Others have viewed these associated clusters of technologies as 'technological styles', or as ideal types of productive organisation, that changed in response to the emergence of new and potentially all-pervasive factors of production such as steel, oil and electricity. Examples of such technological styles were the water style, the steel and electricity style, and the Fordist style (Tylecote 1991:36-53).

A related idea is that long waves are a reflection of the life-cycle of new products of basic innovations. But, in response, it was not clear why the time-span of such life-cycles should always approximate to around half a century. It could be argued that many product life-cycles were very short, perhaps even becoming shorter over time, as technology becomes ever more prevalent within many sectors of the advanced capitalist economy. Sometimes associated with the innovation life-cycle idea was the importance of infrastructural investment to generating long-run growth or stagnation. In this view the alternation of capital-hungry and capital-satiated periods underlay the long wave pattern, with large infrastructural investment (such as industrial and transport complexes) having key significance (van Duijn 1983:138-39). This approach was a direct descendent of the type of explanation first outlined by Kondratiev.

Another approach to understanding long waves emphasized the connection between wars and economic development. From this perspective, the upswing phase of the cycle supported a sustained rise in production, which in turn supported an increase in the potential for war amongst the great powers of the day. A related factor was that the long wave upturn sharpened the potential for conflict over new markets and over new colonial conquests, thus increasing the possibility of war. Once war had broken out

between nation states this had a serious effect on economic growth, tending to dampen the long wave upswing through inflation and the physical destruction of capital (Goldstein 1988:260-74).

However, war could also generate a rise in the application of new technology to military production and also in relevant scientific innovation, which then took some while to feed through into applications in the civilian sectors of the economy, thus laying the foundations for the next long wave upswing.

One factor only rarely considered by long wave theorists was the various psychological factors that might be involved in the cycle propagation process. Periods of the rising wave of the long cycle might be characterized as exhibiting the mass psychology of optimism, which (once first instigated) developed an internal logic of its own, leading to excessively positive evaluations of future business prospects. A glut of risk taking by entrepreneurs usually followed, which appeared (for a time) to justify itself through increased prosperity. However the increased demand for capital, labour and raw materials that inevitably accompanied prosperity eventually resulted in significant price rises, which in turn impacted upon profits and hence upon subsequent evaluations of future prospects (Stoken 1993:84-86). It then only took a very specific problem (for example a limited market crash or a series of bankruptcies) to function as the proximate cause that sparked the shift from the rising wave phase to the declining wave of the long cycle.

Clearly in debt to conventional business cycle theory, this psychological account fitted certain periods of long wave prosperity (such as the post-World War 2 boom) quite well, but whether the over-reactive psychology itself was the underlying

determinate factor at work in cycle generation could easily be disputed. It is probably more accurate to conceive of changes in mass psychology as functioning in mutual interaction with the many others factors involved in secular development, such as technological change and military conflict. Most long wave theorists have relegated psychological factors to only a minor role in their general explanatory scheme.

Finally, an institutionalist approach to theorising long waves has been proposed by David Gordon, under the heading of the social structures of accumulation (Gordon 1978). Within this framework, capitalism was seen as evolving through long waves of growth and decline that were generated by sets of institutions—such as specific corporate structures or systems of labour management - that came into operation and then dissipated. For example Gordon delineated that the post-World War 2 long wave upswing was based on three hegemonic institutional arrangements as follows: 1) Pax Americana; 2) a capital-labour accord; 3) a state-citizen agreement. These institutions began to fail in the early 1970s with challenges from the Third World to Pax Americana, and they were also undermined by domestic forces which opposed the labour and citizens accords from within the USA. In Gordon's approach, economic factors were inextricably intertwined with social and political changes that were on-going within the institutional framework of capitalism, and hence long waves were not by any means an exclusively economic phenomenon. (See O'Hara 2006).

### **Criticisms of the Long Wave Concept**

One of the basic criticisms encountered is that long cycles are of such extended duration that the number of repetitions

required in order firmly to establish their existence as genuine cycles is nowhere near observable with the historical datasets currently available. Even within the long wave framework, it might be asked why the diffusion of technological change (Schumpeterians) or the progression of large capital investments (Jay Forrester) must generate wave-like patterns. Why are the resultant wave patterns periodic? One possible answer to this question was provided by Kondratiev's colleague within the Conjecture Institute, E.E. Slutsky, in a famous article from 1927 that investigated the random causes of business cycles. Slutsky suggested that the summation of random events could generate sine-wave-like phenomena, at least for a definite period of time, although few long wave theorists have made this connection directly.

Some investigators who have tested aspects of the long cycle hypothesis empirically have rejected Schumpeter's idea of regular innovation clusters during the depression phase of the cycle, and have also questioned whether a 'swarming' effect can be detected empirically (Solomou 1988:100). Others have suggested that the long waves in question were actually not in any way periodic and hence the focus on testing for approximate regularity was misguided.

Another line of criticism related to both the data and the methodology employed by Kondratiev and Schumpeter to detect long cycles. It has been pointed out that price and monetary data were often dominant in the analyses of these two pioneers, as opposed to production data. This might suggest a parallel with an implication made by Irving Fisher with respect to the standard business cycle, that the results were simply a long 'dance of the dollar'. Regarding the statistical methodology that was employed,



Kondratiev used ordinary least squares and moving averages, but these procedures assumed that the cyclical movements being uncovered occurred around an equilibrium path, within an unchanging economic structure (Solomou 1988:4,14). Such implicit assumptions were questionable, especially when the period being investigated was so protracted.

Another aspect of the problem was that the statistical manipulation of data by means of a moving average process could generate cyclical swings in a time series when none had existed originally, as had been suggested by Slutsky. But Kondratiev did not even acknowledge this possibility in any way, which was surprising, as Kondratiev had personally requested that Slutsky should become a consultant in the Conjecture Institute in 1926. Finally, some early critics of Kondratiev's methodology pointed out that the trend estimates used were very sensitive to the extension of the data sets employed, and that when additional data was included in the calculations, the timing and amplitude of the cycles that were detected altered considerably (Reijnders 1990:24-25). In general it would not be inaccurate to state that there is some disagreement even among long wave advocates as to the correct methodology to be employed when detecting the secular patterns involved.

### **Policy Relevance**

Long waves have a significant number of policy-relevant consequences that have at least partially been recognised in the literature over a long period of time. Kondratiev was concerned to employ long cycle analysis as a tool for economic forecasting in relation to scheduling the export of primary produce, as part of a market-led industrialisation strategy for the USSR that he developed in the mid-1920s

(Barnett 1998a). Kondratiev's market-based strategy fundamentally contradicted Stalin's agrarian policies after 1929, and Kondratiev was executed as a direct consequence of this conflict. Some later commentators assumed that Kondratiev had been so dispatched because of the implications of his long cycle approach to capitalist development, but in fact his real crime was to tolerate rural differentiation.

With regards to the innovation approach, some more recent authors have emphasised that new groups of firms have periodically broken through to the top echelon of companies based upon their competence in the emerging new technologies that drive long waves. Consequently various specific regimes of regulation, or patterns of coordination and control, were seen as crucial for determining the success or otherwise of particular national and local economies in the context of the world economic system. Institutions such as the IMF, the WTO and the World Bank have often attempted to liberalise markets in terms of facilitating international access, but the uneven development of the world economy, and the uneven spread of new technology, has frequently generated major political and economic crises (Freeman and Louçã 2001:350-70). International bodies thus must take into account the logic of long waves when designing appropriate policies for encouraging growth and development, and governments of less developed countries must be fully aware of the need for encouraging the use of new technology in business enterprise.

Another significant policy consequence of long waves links to political and social conflicts such as strikes and wars. According to some historians, the clustering of social conflict frequently occurred around the turning points of the long waves (Hobsbawn

1964:148). In such accounts it was suggested that outbreaks of strikes tend to cluster near the upper turning point of long waves, as a consequence of workers' movements becoming more powerful after a long phase of industrial expansion. At the lower turning point of the long wave, structural crises of adjustment led to greater job insecurity and hence to a partial quelling of union militancy.

Still others have tried to find regularities relating to increasing and decreasing overall economic inequality within the long wave framework. For example a tendency to reduced inequality in a short period at the end of the downswing and the beginning of the upswing, together with increased inequality concentrated in the early phase of the downswing, has been asserted as existing within many capitalist economies (Tylecote 1991:130). This connects to the issue of the link between long waves and centre-periphery dynamics.

Theorists of the structural under-development of the periphery (Latin America, Africa, Asia) within the capitalist world system have highlighted factors such as the relocation of less profitable production sectors from the core to the semi-periphery, as a mechanism to assist in the launching of a long wave upswing. Immanuel Wallerstein articulated a three-tier model of core, semi-periphery and periphery linked through a world market mechanism, and argued that countries could enter or exist from the semi-periphery in periods of transition, i.e. in periods of long wave crisis (Wallerstein 1991:130). Others have suggested that an inverse relationship between the long wave dynamics prevalent in the capitalist centre against those seen in the under-developed periphery existed at any given time. Certainly the post-World War 2 long wave upswing was more obviously a

phenomenon of the advanced core of nations (Europe and America) than that of the developing world.

Finally, a link between economic long waves and long-run changes in favoured government policy has been highlighted. For example Keynesian interventionism was the dominant ideological convention among many Western governments in the 1950s and the 1960s, i.e. within a period of long wave upturn. With the onset of crisis in the 1970s a definite shift towards the free market ideology of monetarism was observed, at least within the UK and the USA. Marxists would see the latter as one of the ways in which the dominant forces within capitalism were able to partially overcome the crisis through increasing the overall rate of profit, a process which might also function to enable the launching of a new long wave upturn. Whether such changes in government policy were seen as being endogenous to the long wave depended upon the particular theory being employed.

### **Current Issues**

In the most recent period, the long wave concept has been allied to the idea of the emergence of a whole new type of economy driven primarily by knowledge-based activities, the application of new media technologies such as the internet across many industrial sectors, and the increasing weight of the service sector in advanced capitalist economies (Flew 2005: 42). The hypothesis of a fifth long wave beginning at the very end of the twentieth century has been readily asserted, in association with the idea of the birth of a network society based on global information flows appropriate to the post-Fordist age. This new economy was based on information and communications technologies (for example the interlinked microprocessor) and was defined as being

‘weightless’, that is it had non-material output (code, design and information) as its key driving force. It also involved a significantly increased speed of communication facilitated by the new technologies involved. Wallerstein for example has suggested that a post-millennial long wave upswing might be led by a Japanese-American consortium, that would possess a strong hold on both the entrepreneurial and R&D expertise required for success within the new network economy (Wallerstein 1991:131).

However such assertions of immediate novelty tended to ignore the fact that information flows were seen as crucial in the Austrian view of market exchange from as far back as World War 2 (Hayek 1945), and that the evolution of legal and financial institutions as ‘habits of thought’ still exerted a large influence on the particular forms of the new economy that were documented as coming into existence. Another way of approaching this aspect of the debate has been to stress that every long cycle since the end of the eighteenth century has brought with it a ‘new economy’ in terms of the technological revolution(s) that accompanied it, comparisons between the effects of the invention of the telegraph and the internet now being commonplace. Hence there was nothing really new about the early twenty-first century form of the ‘new economy’ that was uniquely relevant to long wave theory.

A further connection that has recently been made is to globalisation, and whether there are any changes that have been brought to the long wave pattern of development through increased economic integration. Since the long cycle idea was always conceived of as being international in scope, heightened global links might only strengthen the regularities that were posited

as being involved, especially if changes in favoured government policy (intervention against the minimalist state) and the evolving social structures of accumulation were accepted as being part of the long wave framework. On the other hand, the development of new political movements in less-developed countries, and also the anti-globalisation movement within developed states, might be seen to challenge and disrupt the pattern associated with the global trade networks of advanced capitalism. In general, the increased complexity associated with globalisation might reasonably be predicted to impact significantly on all existing economic regularities, including those of long waves.

Some more controversial observations for the future have been made by a few long wave adherents. A new long wave based on neural technology and the more extensive use of biotechnology has been predicted for the first half of the twenty first century, a forecast partially at odds with the previously discussed fifth long wave based upon the network society. Biochips and nanotechnology will (it is argued from this perspective) form the basis of the rising wave of this new long cycle through deployment in the healthcare and entertainment sectors of the economy, with existing communications technology becoming allied to the fruits of the maturing biotech revolution.

Whether such a predicted new biotechnology boom will actually provide the underpinnings of a twenty-five-year upturn in classic long cycle form has yet to be seen. Perhaps a more plausibly scenario is that both the network economy and new uses of biotechnology will develop in tandem, and that the ever increasing computerisation of everyday life will act to prolong (and perhaps even merge together) various

technological changes that were in progress at various levels at the end of the twentieth century. Undoubtedly such technological change will continue to impact upon secular economic development in a dramatic and ongoing fashion, and one significant and lasting way of theorising this impact has been through the framework of long waves of growth and development.

## Conclusion

The hypothesis of the long wave has certainly proved to be a fertile and lasting one in terms of the quantity of debate that it has generated and the amount of attention that has been devoted to proving or disproving its existence. Along the way many cycle-related concepts have been clarified and various enlightening empirical observations have been documented. Whether long cycles with a strict periodicity do actually exist or not depends at least in part on how they are initially defined and competing characterisations are rife in the voluminous literature on this fascinating and controversial topic.

## Selected References

- Barnett, Vincent. (2001) "Tugan-Baranovsky as a Pioneer of Trade Cycle Analysis". *Journal of the History of Economic Thought*, Volume 23, Number 4, pp.443-66.
- Barnett, Vincent. (1998a) *Kondratiev and the Dynamics of Economic Development: Long Cycles and Industrial Growth in Historical Context*. London: Macmillan.
- Barnett, Vincent. (1998b) "Dating the Long Cycle Turning Points: Kondratiev and After", in N. Makasheva, W. Samuels and V. Barnett (Editors), *The Works of Nikolai D. Kondratiev*. London: Pickering and Chatto, Volume 1, pp.xxxv-l.
- Berry, Brian. (1991) *Long-Wave Rhythms in Economic Development and Political Behavior*. Baltimore: Johns Hopkins University Press.
- Burns, Arthur and Wesley Mitchell. (1946) *Measuring Business Cycles*. New York: National Bureau of Economic Research.
- Duijn, J.J. van. (1983) *The Long Wave in Economic Life*. London: George Allen.
- Flew, Terry. (2005) *New Media: An Introduction*. Oxford: Oxford University Press.
- Freeman, Chris and Francisco Louçã. (2001) *As Time Goes By: From the Industrial Revolutions to the Information Revolution*. Oxford: Oxford University Press.
- Friedman M., and A.J. Schwartz. (1963) *A Monetary History of the United States, 1867-1960*. Princeton: Princeton University Press.
- Goldstein, J.S. (1988) *Long Cycles: Prosperity and War in the Modern Age*. New Haven, CT: Yale University Press.
- Gordon, David (1978) "Up and Down the Long Roller Coaster", in *U.S. Capitalism in Crisis*. New York: URPE.
- Hayek, F.A. (1945) *Individualism and Economic Order*. London: RKP.
- Hobsbawm, E.J. (1964) "Economic Fluctuations and Some Social Movements since 1800", in Eric Hobsbawm (Editor), *Labouring Men*. London: Weidenfeld and Nicolson.
- Kondratieff, Nikolai. (1925) "Bol'shie tsikly kon'yunktury", *Voprosy kon'yunktury*, Number 1.
- Kuznets, Simon. (1967) *Secular Movements in Production and Prices*. New York, Kelley.
- Makasheva, N., W. Samuels and V. Barnett. (1998) (Editors) *The Works of Nikolai D. Kondratiev*. London: Pickering and Chatto.

- Mandel, Ernest. (1976) *Late Capitalism*. London: Verso.
- O'Hara, Phillip Anthony. (2006) *Growth and Development in the Global Political Economy: Social Structures of Accumulation and Modes of Regulation*. London & New York: Routledge.
- Rainov, T.I. (1929) "Wave-like Fluctuations of Creative Productivity in the Development of West-European Physics in the 18<sup>th</sup> and 19<sup>th</sup> Centuries", *ISIS*, Volume 12 Number 37.
- Reijnders, Jan. (1990) *Long Waves in Economic Development*. Cheltenham, UK: Edward Elgar.
- Rostow, W.W. (1990) *Theorists of Economic Growth from David Hume to the Present*. Oxford: Oxford University Press.
- Schumpeter, Joseph. (1939) *Business Cycles*. New York: McGraw-Hill.
- Solomou, Solomos. (1988) *Phases of Economic Growth, 1850-1973*. Cambridge: Cambridge University Press.
- Stoken, Dick. (1993) *The Great Cycle: Predicting and Profiting from Crowd Behavior, the Kondratieff Wave and Long-Term Cycles*. Chicago: Probus Publishing.
- Tylecote, Andrew. (1991) *The Long Wave in the World Economy*. London & New York: Routledge.
- Wallerstein, Immanuel. (1991) *Geopolitics and Geoculture: Essays on the Changing World-System*. Cambridge: Cambridge University Press.
- Zarnowitz, Victor. (1992) *Business Cycles*. Chicago: NBER.

Vincent Barnett  
 Centre for Russian & East European Studies  
 University of Birmingham  
 Birmingham, UK  
[vincentbarnett@postmaster.co.uk](mailto:vincentbarnett@postmaster.co.uk)

## Moral Hazard and Adverse Selection

Jong Shin Wei

### Introduction

Risk-pooling is where insurance is conceived. Imagine that it is believed that one out of 1000 cars will be stolen each year, and if each car owner (as an insured) contributes to a fund (managed by an insurer) 1 percent of the present value of his car, the sum will be large enough to reimburse the poor guy whose car gets stolen. For the rest of owners, misfortune does not strike, meaning that they pay the premium and get nothing yet they are contented. However, a car owner might do better than break-even when misfortune strikes. In that case, the insured has the incentive to expend less effort in protecting his car and might even violate moral code to have the stealing take place. Such tendency of changing behavior after purchasing insurance is referred to as *moral hazard*. On the other hand, people who buy insurance often have a better idea of the risks they are facing than the sellers do. People who know that they face large risks (e.g., having the necessity of regularly parking in the rough neighborhood; often forgetting to lock) are more likely to buy auto-theft insurance than people who face small risks. Being unable to sort out potential buyers of insurance by their true types (or characteristics) creates the problem of *adverse selection*.

Each transaction has two sides: supply and demand. If both sides possess unequal amount of information concerning the (quality of) objects in trade, *informational asymmetry* follows. In the example of moral hazard, we use *hidden actions* to account for the unverifiable actions of the insured which increase the probability of car theft. Meanwhile we use *hidden characteristics* to

explain the inability of separating buyers of low-risk from those of high-risk.

Adverse selection was a long-time concern in the insurance industry. It referred to the tendency of “identifying or finalizing” insurance-buyers through a non-random selection from the population in which only those who expect to have the highest expected claims will buy insurance. This term now has a much richer meaning and ampler applications, particularly in the contractual relation involving principals and agents. Adverse selection refers to the kind of *precontractual opportunism* that arises (1) when one party to a bargain has (by nature) private information about something that affects the other party’s net benefit from the contract and (2) when only a subset of those endowed with superior information will accept the contract and subsequently jeopardize the other side of the contract.

Moral hazard originally also was an insurance term referring to the unverifiable tendency of people with insurance to reduce the care they take to avoid or reduce the incurred losses (covered by such insurance policy). Now by moral hazard we mean the form of *postcontractual opportunism* that arises when actions of one party required or desired (in spirit) under the contract are not freely observable (by the other side), creating a self-interest incentive to take unusual risks. Hence they tend to jeopardizing the other party’s (anticipated) net benefit from the contract.

Both modern definitions, together with many real-life illustrations, can be found in Milgrom and Roberts (1992), for instance.

### A Classic Example

Here is an example modified from Akerlof (1970) that sheds light on *market failure* (which is the extreme consequence of the

problem of adverse selection) resulting from the presence of informational asymmetry.

The set of used cars is indexed by a quality parameter  $q$ , which is uniformly distributed between 0 and 1. Assume that the number of potential buyers exceeds the number of sellers, partially justifying the following assumptions on reservation prices. For a car of quality  $q$ , the reservation price of a buyer (i.e., the highest price a buyer can bear) is  $1.5q$ ; the reservation price of a seller (i.e., the lowest price for a seller to stand) is  $q$ . Obviously, the gap between  $1.5q$  and  $q$  marks the so-called “room for bargaining”. In an ideal classic “perfect competition” framework *a la* Arrow-Debreu, with quality of cars assumed to be known to both sides, all cars will be sold at the equilibrium price being  $1.5q$  for a car of quality  $q$ . Nevertheless, Akerlof took the lead in assuming that sellers know the quality of cars (to be sold) while buyers don’t. We find it very convincing as used car is a good which can vary in quality and whose quality is known only to the owner. In light of seller’s reservation price, taking any car that is for sale at price  $p$ , buyers know that its quality cannot exceed  $p$ . From the closed interval between 0 and  $p$ , buyers infer that the average quality of any car for sale at price  $p$  is  $0.5p$ . So we see the informational asymmetry in the sense that sellers know all while buyers know only the average quality of cars (thru the information conveyed by price tags). Since the average quality is  $0.5p$ , the buyer’s reservation price now becomes  $(1.5)(0.5p) = 0.75p$ , which is less than  $p$ , implying that any car of true quality  $p$  will not be traded. With help from this oversimplified example, we see how informational asymmetry might interfere with the effective operation of a perfectly competitive market. Or put in strong words, we just demonstrate a market failure.

What we analyzed above is the most extreme consequence of the problem of adverse selection, yet generally speaking, we are much interested in seeing that cars of low quality are traded while cars of high quality are not. Such phenomenon is under the category of adverse selection. To see it, envisage that a mandatory one-price rule requiring all used cars to be traded at some price, say 0.5 (as the average quality), takes effect. No car of quality index exceeding 0.5 will be offered by sellers at that price! What buyers get in the market will be cars of quality index no greater than 0.5. The term adverse selection captures the idea that the selection of used cars offered in the market is determined in a way that is adverse to the interest of poorly informed buyers.

### **Hidden Characteristics and Actions**

A non-technical sketch on the intellectual development of these concepts is made for policy-oriented readers. Back to the 1950s, Leonid Hurwicz originally from Poland, who is one of the most influential contemporary micro-economists, did the pioneering work on mechanism design and informational efficiency. Each mechanism is just like a production process by which economies are “injected” as inputs and allocations of resources come out as outputs. The operation of any mechanism requires economic agents to report their privately known true characteristics, which are prescribed tasks to be carried out by agents. However the reported economy need not be the true economy as some agents might have the incentive to lie. We see unilateral manipulation via misrepresenting preferences (even in the absence of public goods), withholding or destroying endowments, overstating productivity (or exaggerating human capital), among many others, blooming before 1980s. Incentive

compatibility requires all agents truthfully reveal true characteristics. Impossibility theorems abound, which roughly say that no mechanism can simultaneously satisfy Pareto efficiency, individual rationality (or voluntary participation constraint), and incentive compatibility. Besides the incentive to misrepresent characteristics (including valuations), a second sort of incentive problem arising from precontractual informational asymmetry is known as adverse selection. Once we are in the postcontractual stage, actions may not be observable while output levels are. A real estate manager may work hard or shirk. His effort may not be correctly inferred from the sales record as the demand side is often subject to some random factors. From hidden actions on his part we see moral hazard.

Real-life examples of hidden characteristics and hidden actions are numerous and often coexisting. For example, consider an extended warranty offered for a blender. Those who expect to use this product with less caution or may occasionally push the appliance to the limit would like to pay for this special offer. Hidden characteristics talk. Once it is paid, hidden actions matter as the owner may break it simply because of his not reading the manual carefully.

A good example of hidden characteristics is childbearing plans—the time table of having children particularly for career women. If an insurance company were to issue an individual health insurance policy that covers the medical costs associated with pregnancy and delivery, soon the policy would be purchased disproportionately by women planning to bear children in the very near future. (Once the policy takes effect, some women may exercise less caution in birth-control, hence creating moral hazard.)

With such severe adverse selection it is not surprising that such policy is no longer available in the U.S. In many other countries, the inadequacy of private health insurance has led government to nationalize the provision of health care, at which adverse selection problems are replaced by moral hazard. In the U.S., group health insurance, such as employer-provided medical insurance plans, having the pregnancy and delivery benefits packaged with some other health-care benefits, might alleviate adverse selection problems to some extent.

### **Adverse Selection: Problems Identified and Resolutions**

Economists cannot talk about adverse selection without quoting *Gresham's law*. As Ritter and Silber (1993) put it, Sir Thomas Gresham, financial adviser to Queen Elizabeth I, is said to have coined the celebrated phrase “*bad money drives good money out of circulation*,” meaning that if two types of money of the same denomination serve as media of exchange, with one containing less valuable (or debased) metal and the other containing more valuable metal, then the former shall remain in circulation yet the latter will be hoarded. This is because that everyone will try to hold on to the more valuable and pass on the less valuable.

Rothschild and Stiglitz (1976) brought the notion of adverse selection to a perfectly competitive health insurance market. They showed that (1) only a separating equilibrium can exist where healthier individuals either under-insure or do not insure at all; (2) with sufficiently large proportion of sicker individuals, no equilibrium will exist. The former demonstrates adverse selection; the latter resembles market failure.



Can *hidden* characteristics (responsible for adverse selection) be made known or discovered by expending effort and money? It depends. For instance, self-serving claims such as “I am the best candidate you can possibly get from the market” are most unlikely to be credible; conducting a thorough background check on new employees is extremely costly, to say the least. Yet signaling and screening might help. In *signaling*, the privately informed party can take the lead in adopting behavior (such as communicating) in a way that hidden characteristics can be revealed to the uninformed party. According to Frank (2000), for a signal to be effective, it must (1) be costly-to-fake and (2) satisfy the full-disclosure principle. The principle says that if some informed individuals stand to benefit by revealing a favorable value of some characteristic, others will be self-driven to disclose their even less favorable values (in order not to be identified as those having the lowest value). According to Spence (1973), educational attainment, in addition to building human capital, can be used by job-seekers to signal productivity. Offering product warranties and building a brand name are also examples of signaling, so is making a deep first impression by taking whom you found thru the romance ads in the New York Times to a fancy restaurant on the first date.

While sellers signal, buyers screen. *Screening* refers to activities undertaken by the less informed party in order to separate different types of the informed party along some dimension. In short, screening tries to sort bad information from good information.

From the idea of screening, a firm can offer a menu of contracts to salespeople. For example, salespeople who know the territories well and are highly motivated will select high-commission but low-salary

contracts; salespeople with less skill, inadequately motivated, or lack of self-confidence will opt for low-commission with median-salary contracts and face less risk.

Note that, to date, there appears to be no consensus about the presence of adverse selection as far as empirical evidence is concerned. For instance, while Marquis and Phelps (1987) found evidence of adverse selection in the supplemental insurance market, Dowd et al (1991) reported an opposite finding.

### **Moral Hazard: Problems Identified and Resolutions**

A synthesis on prospects and technicality of home equity conversion in various forms, including potential problems and remedies of moral hazard, can be found in Shiller and Weiss (1998). By home equity conversion in various forms we mean several plans that enable homeowners to convert their illiquid and risky investments in their own homes to some other uses, hence reducing their exposure to real estate risk. Examples of these risk-reducing institutions include reverse mortgages and home equity insurance. The former are contracts providing regular payments and/or a lump sum to the homeowner in exchange of such debt being repaid only when the home is sold, the owner no longer resides at premise, or the owner dies. Clearly this involves some partial home price risk sharing for homeowners on one hand; on the other hand, homeowners might fail to take necessary steps to maintain the value of their homes, known as the problem of moral hazard. As to the resolution, they recommend indexing the debt to the real estate price index so as to penalize homeowners for deviations of the selling price from the value predicted by the index as well as the original selling price. The latter, still in the proposal stage but

might see the light soon, refers to a policy that insures the price of a home on resale. To eliminate homeowners' moral hazard, it suffices to specify that only the decline in the real estate price index for the region and kind of home, *not* the decline in the resale price of the home itself, will be covered. Four other home equity conversion forms along with a calibrated model for analyzing moral hazard are documented in Schiller and Weiss (1998).

Two well-known techniques can be employed to discourage moral hazard in insurance markets: deductible and co-payment. A *deductible provision* offers the insured a choice of the amount of each claim to leave uncovered. In other words, in any damage claim the insured must pay the damage up to that preset limit. A *co-payment provision* specifies the percentage of the bill (at each claim) to be picked up by the insured. Of course, the larger deductible provisions or co-payments are, the more policyholders save in premiums. If moral hazard is of the type that increasing risk of a loss is concerned by the insurer, then insurer can benefit from deductible provision due to (1) saving the trouble and expenses in processing numerous small claims and (2) encouraging greater care. If the nature of moral hazard has much to do with size-increase in the loss, then co-payment provision should be preferred by the insurer.

### Exercise Cautions in Policy Debates

Is moral hazard a devil created by the system or could it simply be an unavoidable consequence of the system? To make clear what unavoidable consequence means, we can think about the law requiring helmet-wearing whenever one is riding on a motorcycle. Many will argue that we might see more speedy and reckless riding once it is implemented. Should such law be banned,

riders without wearing helmets would surely be somewhat more careful because that now they realize a sudden slip in high speed could damage the brain, hence very likely take life away. Yet few will advocate banning it. Why? Wearing helmets does save lives and its benefits obviously outweigh any adverse impact on driving safety. So, we treat occasional observations of reckless riding as an unavoidable consequence, not moral hazard, of mandatory requirement of helmet-wearing. Having this analogy in mind, two International Monetary Fund (or IMF for short) economists, Lane and Phillips (2000) argued that (1) moral hazard might just be an unavoidable consequence of the operation of IMF and (2) empirical evidence (by examining the interest rate spreads paid on emerging market bonds during 1995-1999) does not seem to support the well-spread critics, naming that *the knowledge* that IMF financing (or "massive bailouts") will be made available in the event of a financial crisis (e.g., in Mexico, East Asia, Russia, and Brazil) *ironically makes the crisis more likely to occur*.

### Selected References

- Akerlof, G. (1970) "The Market for Lemons: Qualitative Uncertainty and the Price Mechanism", *Quarterly Journal of Economics*, Volume 84, pp. 488-500.
- Dowd, B., R. Feldman, S. Cassou, and M. Finch. (1991) "Health Plan Choice and the Utilization of Health Care Services", *Review of Economics and Statistics*, Volume 77, pp. 85-93.
- Frank, R.H. (2000) *Microeconomics and Behavior*. 4th Edition. New York: McGraw-Hill.
- Lane, T. and S. Phillips. (2000) "Does IMF Financing Result in Moral Hazard?", *IMF Working Paper, Number 168*. Washington DC; IMF.

- Marquis, S. and C. Phelps. (1987) "Price Elasticities and Adverse Selection in the Demand for Supplemental Insurance", *Economic Inquiry*, Volume 25, pp.299-313.
- Milgrom, P. and J. Roberts. (1992) *Economics, Organization and Management*. Englewood Cliffs, NJ: Prentice-Hall.
- Ritter, L.S. and W.L. Silber. (1993) *Principles of Money, Banking, and Financial Markets*. 8th edition. New York: BasicBooks.
- Rothschild, M. and J.E. Stiglitz. (1976) "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information", *Quarterly Journal of Economics*, Volume 90, pp.629-649.
- Shiller, R.J. and A.N. Weiss. (1998) "Moral Hazard in Home Equity Conversion", paper presented at AREUEA-ASSA session at Chicago, January 1998.
- Spence, A. M. (1973) *Market Signalling: Information Transfer in Hiring and Related Processes*. Boston, MA: Harvard University Press.

*Jong-Shin Wei*

*Department of International Business  
Administration, Wenzao Ursuline College  
of Languages, Taiwan  
93001@mail.wtuc.edu.tw*

## National Accounts

Yngve Abrahamsen and Jochen Hartwig

### Introduction

“National income estimates date back to the latter seventeenth century. But it has been only in the past half century or so since World War II that economic accounts have developed in their present form, becoming an indispensable tool for macroeconomic analysis, projections, and policy formulation” (Kendrick 1996:1). This article reviews the historical stages of national income accounting that paved the way to the latest version of the United Nations’ System of National Accounts (SNA 1993), which is currently used by statistical offices throughout the world. The design of the *SNA 1993* is then presented in some detail, and with the aid of a numerical example. Finally, the links between national accounts and governance issues are established.

### Historical Development

William Petty (1623-1687), whom Marx lauded as the ‘father of Political Economy, and who to some extent was also the founder of Statistics’, was the first to provide rough estimates of ‘national income’ in his *Political Arithmetick*, that appeared in print posthumously in 1690. This remarkable work combines already, and in a perplexing manner, most of the elements that have been regarded as crucial for national accounting up to the present day. Not only does Petty acknowledge that ‘The Labour of the People’ is the source of national income, which is echoed in modern production accounts (see below), but he also estimates the division of national income between wages, rents, interest, and profit; and he opposes this with the disposition of income by giving an estimate of annual domestic

consumption expenses. As we will see below, Petty’s practice comes remarkably close to the modern system that establishes the link between the ‘distribution of income account(s)’ and the ‘use of income account(s)’ by means of double-entry bookkeeping. Finally, Petty also appears to be ‘modern’ with respect to the intentions behind his national income calculations. Above all, he aims at devising a means to compare the relative strength of the English economy with that of her main commercial rivals of the time, that is, France and the Netherlands. To allow for international comparisons is still an important task of present-day national accounts. Also, governance aspects of national accounting have not gone unnoticed by Petty, as is evidenced by the title of the fifth chapter of the *Political Arithmetick*: ‘That the Impediments of England’s Greatness, are but contingent and removable’ (cf. Hartwig 2001).

For the next two hundred years, progress in national accounting was slow. Admittedly, François Quesnay’s *Tableau Economique* (of 1766) which, for the first time, envisaged exchanges in an economy as a circular flow was a precursor of later input-output-tables that now form a part of national accounts. Also, there was an important contribution coming from Adam Smith who, in *The Wealth of Nations* (1776), laid emphasis on productive activities that ‘fix themselves’ in commodities rather than services.

This concept was later modified by Karl Marx, who included the provision of services as being productive, as long as they are organised along capitalist lines and thus yield surplus value. A modified ‘Marxian’ system became the basis of the ‘Material Product System’ of national accounts prevalent in the Soviet Union and other communist countries—even in France, for

some time. Alfred Marshall included services in his system of accounting. The production and services view of accounts was adopted by the United Nations in their recommendations for compiling national accounts.

Up to 1890, national income estimates were made in France, Russia, the United States and Austria, adding to the English estimates. All these estimations, though, were restricted by the poor quality of the data at hand; and they comprised only particular years or short periods of time. The focus of the investigations remained 'national income' as an aggregate. Even Simon Kuznets's (1934) study—which is widely regarded as a pioneering effort—in fact showed the distribution of national income by type of income received, but not yet the breakdown of national product by type of final demand.

Until the end of World War I, national income estimates were prepared in eight additional countries, including Australia, Germany and Japan. Gradually, government agencies were taking over the task of composing statistics from individual researchers, hence data quality improved. Two incidents fostered the final breakthrough of national accounting: first, J. M. Keynes's *General Theory of Employment, Interest and Money* (1936) encouraged thinking in terms of macroeconomic aggregates such as consumption and investment demand. Also, Keynes proposed an appropriate delineation for these aggregates to show that production, distribution and appropriation aspects of national income are in fact inextricably interwoven. But the final impetus for national accounts came from the outbreak of World War II. In urgent need of a reliable basis for their war budgets, the British government advised economists at the

Central Statistical Office to prepare a set of income and expenditure estimates. These estimations—published as an appendix to the 1941 U.K. budget—were made by James Meade and especially Richard Stone. Both Stone and Meade were later awarded the Nobel Price in Economics; the former explicitly 'for having made fundamental contributions to the development of systems of national accounts and hence greatly improved the basis for empirical economic analysis'. Stone's work (1947, 1951) was fundamentally new in that it integrated national income in a double-entry bookkeeping format so that every item entering as income would be matched by some expenditure item on the other side of the account. Such a double-sided account in principle allows for an analysis of the distribution and disposition of national income. Stone also disaggregated the appropriation account by major sectors and linked it to the consolidated production account by suitable double entries. After the war, his approach strongly influenced the first recommendations published by the United Nations' Statistical Office for reporting national income and its components (United Nations 1953).

During the following two decades, virtually all nations began to set up national accounts. On a conceptual level, the thrust was towards improving data quality on the one hand and to broaden the group of economic activities covered by the System of National Accounts on the other. Hence, the revised *SNA 1968* (cf. United Nations, 1968) integrated the production, distribution and appropriation accounts with financial accounts (or flow-of-funds accounts, as they are called in the U.S.). Also, the production account was disaggregated into input-output tables in respect of industries and commodities. The inclusion of balance sheet

accounts was intended to secure full stock-flow-consistency of the respective aggregates over time. Constant-price data for the supply and disposition of goods and services were also integrated. Admittedly, not all countries saw themselves able to devote enough resources to their statistical offices to arrive at the intended high degree of sophistication; and still others, like the U.S., preferred sticking to their own national income accounting rules that differed somewhat from the international standards. Nevertheless, in 1993, yet another version of the SNA was published (cf. Inter-Secretariat Working Group on National Accounts 1993); a new version that—although it did not introduce any major conceptual changes—once again raised the level of detail and thus the complexity of the system as a whole. At the time this entry was written, even several developed OECD countries, such as New Zealand, Switzerland, and Turkey, have not yet—or only recently—managed to switch their accounting standards to the *SNA 1993*. The next section describes this *SNA 1993* in detail, particularly with a view to the expansions vis-à-vis the *SNA 1968*. (For a more detailed exposition of the history of national accounts, cf. Studenski 1958, 1961, and Kendrick 1970.)

## **The System of National Accounts 1993**

### *Accounting Principles*

Most economic activities have a two-sided nature. For example, for each sale there must be a purchaser; and for each credit there must be a debtor. Accounts are a convenient way of making allowance for this two-sided nature by recording the two relevant perspectives on economic activities on separate sides—the left-hand side and the right-hand side—of the account and by

postulating that these two sides will have to balance.

The *SNA 1993* records two basic forms of economic entities: flows and stocks. Flows result from economic actions within a given period of time while stocks are nothing but the accumulation of prior flows reported at a certain point of time. The *SNA 1993* distinguishes two kinds of flows: transactions and ‘other flows’. Transactions are based on mutual agreement of two separate parties to exchange something. The accounts record the *values* of goods, services or assets involved in the transactions rather than the physical things themselves. Sometimes, though, the *SNA* treats actions that do not involve an exchange like a transaction, e.g., production for one’s own use or public expenditures. In these cases, values have to be imputed for lack of market prices, which sometimes proves difficult in practice. ‘Other flows’ is a category that was introduced in *SNA 1993*. It describes changes in the value of assets or liabilities that are not the result of transactions.

*SNA 1993* consists of three general types of accounts: current accounts, accumulation accounts, and balance sheets. *Current accounts* record transactions related to the generation, distribution, and use of income. Entries on the right-hand side of current accounts are *resources*; this means transactions that add to economic value, whereas entries on the left-hand side are *uses* (that reduce economic value). The balancing item for each account (the excess of resources over uses) is, of course, also on the left side. Since the *SNA 1993* is a system of double-entry bookkeeping, this balancing item has to be offset somewhere. In fact, the balancing item is a resource for the next account in a sequence (as will be shown in the next section).

The last balancing item in the current accounts is ‘saving’—which is a resource for the *Accumulation Accounts*. Accumulation accounts record not only transactions but also ‘other flows’ that add to or reduce assets (on the left) and liabilities (on the right). The balancing item, that is, changes in net worth is also on the right side of accumulation accounts. The *balance sheets*—which had not been fully integrated into earlier versions of the *SNA*—record stocks. There are two balance sheets: an opening balance and a closing balance, both with total assets on the left and total liabilities and net worth on the right side. The difference in net worth between the

opening and the closing balance has to be identical to the sum of the balancing items from the accumulation accounts.

### *The Sequence of Economic Accounts*

This section introduces the *SNA 1993* in the form of a sequence of so-called T-accounts. The (arbitrary) numerical values shown in the accounts are the same as, or derived from, those in *SNA 1993*, pp. 55-56. The alphanumerical codes that are attributed to each transaction as well as sundry subcategories are here suppressed in favour of greater clarity. (For a more detailed presentation, cf. Carson 1996; Jackson 2000.)

#### Production Account

<i>Uses</i>	<i>Resources</i>
Intermediate consumption..... 1,883	Output..... 3,604
Gross domestic product. .... 1,854	Taxes less subsidies on products..... 133
Consumption of fixed capital. .. 222	
Net domestic product..... .. 1,632	
Total..... 3,737	Total..... 3,737

The *Production Account* is the first of the current accounts. Output is the prime resource of the economy. Since, in *SNA 1993*, the preferred valuation for output is at so-called basic prices, taxes less subsidies on products have to be added to arrive at a valuation at market prices. One possible use for output is using it up in the production of further output. This ‘intermediate consumption’ is shown on the left side of the *Production Account*. The item that balances the account is the ‘domestic product’ which is recorded both gross and net of the consumption of fixed capital. In the other current accounts, the balancing items are also expressed in gross as well as in net terms, the difference always being the same. Gross/net domestic product is valued at market prices.

The dual: basic/market prices is not to be conflated with the dual: current/constant prices. For all transactions in goods and services, constant-price data are provided in addition to the values at current prices. The second account is called the *Generation of Income Account*. It is the first in a series of three accounts that deal with the distribution of income, or, put differently, with the allocation of market revenues to the input factors. The *Generation of Income Account* is concerned with the part of the distribution that is directly linked to production. Gross domestic product (GDP) is divided between employees (who receive a compensation for their labour services) and the government which receives taxes (less subsidies) on production and imports. The remaining part (the balancing item) accrues to corporations

as gross operating surplus, or as ‘mixed income’ to households combining labour and entrepreneurship, as in, e.g., unincorporated enterprises. The income accruing from the ‘production’ of housing services also forms part of the ‘gross (net) operating surplus/mixed income’, regardless of whether these services are sold on the

market (to tenants) or whether they are produced for the own use by homeowners. The rental value of owner-occupied residences is an important example for an imputed transaction in the *SNA* for which it may not be easy to arrive at a proper estimate.

### Generation of Income Account

<i>Uses</i>		<i>Resources</i>	
Compensation of employees.....	762	Gross domestic product.....	1,854
Taxes less subsidies on production & imports	191		
Gross operating surplus/mixed income. ....	901		
Consumption of fixed capital .....	222		
Net operating surplus/mixed income.....	679		
Total.....	1,854	Total.....	1,854

Next comes the *Allocation of Primary Income Account*. It records the same kinds of incomes as the *Generation of Income Account*, but it takes the perspective of the recipient rather than the producer. Also, it shows explicitly the income accruing from property, such as interest and dividends, both payable and receivable. We have to keep in mind, though, that the material basis

for property income is always constituted by production. It is not possible, for instance, that the economy as a whole receives income in the form of gross operating surplus and that it receives an additional income in the form of dividends. Rather, the dividends are in a way contained in the gross operating surplus: they are a means to re-distribute this surplus.

### Allocation of Primary Income Account

<i>Uses</i>		<i>Resources</i>	
Property income.....	391	Gross operating surplus/mixed income.....	901
Gross national income.....	1,883	Compensation of employees.....	766
Consumption of fixed capital.....	222	Taxes less subsidies on production and imports.....	191
Net national income.....	1,661	Property income.....	416
Total.....	2,274	Total.....	2,274

Therefore, the entries for ‘property income’ in the above account can only refer to cross-border income streams. Some economies have invested capital abroad and receive property income therefrom. Also, foreigners have invested capital in the domestic economy, hence some part of the gross operating surplus/mixed income goes abroad (here: 391). If we compare the entries for ‘compensation of employees’ in the

*Generation of income account* and the *Allocation of primary income account*, respectively, we notice a difference of 4.

This difference stems from the fact that the *Production Account* and the *Generation of Income Account* follow the so-called ‘domestic concept’: they record productive activity within the borders of a certain nation. In our example, all enterprises producing inside the country pay out 762 to



their employees that live both at home and abroad. These 762 are a reduction in value (use) from the producers' point of view. In the transition to the *Allocation of Primary Income Account*, however, the perspective switches toward the 'national concept' which describes the consequences for the *residents* of a country. (Residents are conventionally understood to be those people who live within the borders of a country for at least one year – irrespective of

their nationality.) Resident households receive as compensation for their labour services 766 from both domestic and foreign employers. These 766 are a resource (addition to value) under the national concept. The SNA is thus a fusion of both the domestic and the national concepts, which will make it necessary to close the system by adding a special account that records all transactions with non-residents (see below).

### Secondary Distribution of Income Account

<i>Uses</i>		<i>Resources</i>	
Current taxes on income, wealth, etc.....	212	Gross national income.....	1,883
Social contributions.....	322	Current taxes on income, wealth, etc.....	213
Social benefits other than social transfers in kind.....	332	Social contributions.....	322
Other current transfers.....	269	Social benefits other than social in transfers in kind.....	332
Gross disposable income.....	1,854	Other current transfers.....	239
Consumption of fixed capital.....	222		
Net disposable income.....	1,632		
Total.....	2,989	Total.....	2,989

The *Secondary Distribution of Income Account* is the last in the sequence of distribution accounts. (In previous versions of the *SNA*, this account was merged with the *Allocation of Primary Income Account*.) The transactions recorded here are unidirectional: residents, for instance, pay taxes for which they do not receive a counterpart in return. Also, the government (including the public social security system) pays out social benefits without receiving anything directly in return. Looking at the numerical entries in the account, we recognise that uses and resources broadly match—the small differences being due to cross-border transactions (that do not have great importance in the field of income redistribution). Let us give an example of how to interpret the entries. The 'social contributions' (of 322) are a 'resource' for the economy as a whole in that they

constitute an addition to economic value for some unit. Likewise, they are a 'use' because they reduce the economic value for those units that have to pay the bill. For social contributions, both these entries exactly match; and this means that there are no cross-border transactions involved here. But this need not be so: there may be cross-border transactions originating from transnational commuters.

Through the back-door, we have introduced the term 'institutional unit', which *SNA* (1993:87) defines as an "economic entity that is capable, in its own right, of owning assets, incurring liabilities and engaging in economic activities and in transactions with other entities". Units with similar functions are grouped together into 'institutional sectors'. In *SNA 1993*, there are five institutional sectors: *Nonfinancial corporations*, *Financial corporations*,

*General government, Households, and Nonprofit institutions serving households.* Note that, although our focus has been on the highest level of aggregation so far, that is, on the economy as a whole, the sequence of accounts described in this section is also advanced for each of the five sectors, thereby creating a much more detailed picture of the economy. In our example, it would turn out that the sector for which

social contributions are a resource is the general government, whereas the sector for which they are a use is the household sector. (In the SNA framework, the employers' share in the social contributions is seen as part of the compensation for labour services and is thus already included in gross salaries.) Of course, the higher the level of aggregation becomes, the more information of this kind will be lost.

#### Use of Disposable Income Account

<i>Uses</i>		<i>Resources</i>	
Final consumption expenditure.....	1,399	Gross disposable income.....	1,854
Adjustment of the change in net equity of households on pensionfunds...	..11	Adjustment of the change in net equity of households on pension funds.....	11
Gross saving.....	455		
Consumption of fixed capital.....	222		
Net saving.....	233		
Total.....	1,865	Total.....	1,865

The last current account is the *Use of Disposable Income Account*. It shows (for each sector) the allocation of disposable income between final consumption expenditure and saving. Note that the sectors *financial* and *nonfinancial corporations* do not consume by definition. The adjustment term for the change in net equity of households on pension funds, which is a new category in SNA (1993), is necessary for the following reason: on the one hand, pension funds are treated as if they were owned collectively by households. This means that flows into and out of these funds should remain internal to the household sector. On the other hand, these payments are included in the social contributions and social benefits, respectively (as recorded in the *Secondary Distribution of Income Account*), because it is believed that to include them accords better with the perception of the households concerned. So, pension contributions and receipts affect disposable income, although they should not, in

principle. The adjustment term takes care of this issue and thus assures that the balance of pension contributions over pension receipts does not enter into household saving.

A notable departure of the *SNA 1993* from its predecessor must be noted at this juncture. It concerns a reallocation of consumption expenditures between governments and households. Earlier, government consumption consisted in the provision of all public services, valued at factor cost—which implies that government consumption was largely identical to the wage bill of public employees. Now, this sum is broken into two parts: (1) expenditures that augment consumption of individuals, and (2) collective consumption (such as, e.g., defence). Public expenditures on education, health, social security and welfare, sport and recreation, and culture—the first category—are thus shifted towards the household sector. This is done in the *Redistribution of income in kind account*, which transfers the *Use of Disposable*

*Income Account to the Use of Adjusted Disposable Income Account.* (Note, that this

transfer is uninteresting at the level of the total economy.)

### Capital Account

<i>Changes in assets</i>		<i>Changes in liabilities</i>	
Gross capital formation.....	414	Net saving.....	233
Consumption of fixed capital.....	-222	Capital transfers receivable.....	62
Acquisitions less disposals on non-produced non-financial assets.....	0	Capital transfers payable.....	-65
		Changes in net worth due to saving and net capital transfers.....	230
Net lending (+)/Net borrowing (-).....	38		

The first of four *Accumulation Accounts* is called the *Capital Account*. It records changes in non-financial assets as a result of *transactions*. Saving, being the balancing item of the current accounts, is the starting element of the accumulation accounts. In the *Capital Account*, net savings plus the balance of capital transfers receivable over capital transfers payable (the latter balance being negative in our example) add to net worth. The increased net worth can be used to finance the (net) accumulation of non-financial assets: fixed capital, inventories and (new to *SNA 1993*) ‘valuables’ like precious metals and stones, jewellery, works of art etc. Non-financial assets that cannot be produced but acquired include intangible assets like patented entities, goodwill, etc.

Also, land and subsoil assets can be bought and sold—not so much by the economy as a whole—but certainly between institutional sectors. In former versions of the *SNA*, non-produced assets have either gone unnoticed or have been recorded as intermediate consumption. The same holds

The balancing item of the *Capital account* is net lending or net borrowing, respectively. If a sector or an economy has a positive entry for net lending, then it does not use up all its financial means for its own physical) capital formation, but provides finance to other sectors or economies. The

good for computer software, which used to be regarded as intermediate consumption but is now recorded as gross capital formation and, if produced, as value adding. Therefore, the gross domestic product under *SNA 1993* is higher *ceteris paribus* as compared to previous versions of the *SNA*. See Figure 1, below:

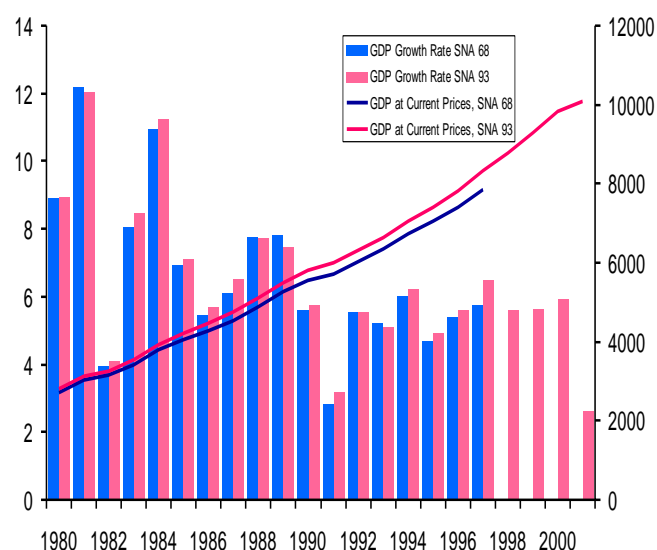


Figure 1: U.S.-GDP at current prices according to *SNA 1968* and *SNA 1993* (Source: *OECD*)

household sector is typically a net lending sector. Also, there are countries that have a long-standing record as net lenders, such as, e.g., Belgium, Japan, and Switzerland, and other countries that are in constant need of borrowing from abroad, such as, e.g., the U.S.

### Financial Account

<i>Changes in assets</i>	<i>Changes in liabilities</i>
Net acquisition of financial assets..... 641	Net incurrence of liabilities..... 603
	Net lending (+)/Net borrowing (-). .... 38

Net lending is also the balancing item for the second in the sequence of accumulation accounts, the *Financial Account*. It shows how the lending and borrowing is carried out via various financial instruments (not shown in the above account) like loans, shares etc.

*SNA 1993*, p. 34, notes: “In principle, net lending or net borrowing is measured identically whichever account is considered. In practice, achieving this identity is one of the most difficult targets of national accounts”.

### Other Changes in Volume of Assets Account

<i>Changes in assets</i>	<i>Changes in liabilities</i>
Economic appearances of assets..... 27	
Natural growth of non-cultivated biological resources.... 4	
Economic disappearance of nonproduced assets -9	
Catastrophic losses & uncompensated seizures. -11	Catastrophic losses & uncompensated seizures 0
Other volume changes nec..... 4	Other volume changes ec... -2
Changes in classification and structure..... 0	Changes in classification & structure..... 0
	Changes in net worth due to other changes in volume of assets..... 17

The *Other changes in volume of assets account* and the *Revaluation account* (below) are innovations of *SNA 1993*. They comprise ‘other flows’ (other than

transactions). Examples for ‘other changes in volume of assets’ are discoveries or depletions of subsoil resources or destruction by natural disasters.

### Revaluation Account

<i>Changes in assets</i>	<i>Changes in liabilities</i>
Nominal holding gains/losses	Nominal holding gains/losses
...On Nonfinancial Assets..... 280	On Liabilities 76
...On Financial Assets..... 84	
	Changes in net worth due to nominal holding gains (+)/losses (-)..... 288

The *Revaluation account* records holding gains or losses due to asset price changes

during the accounting period, both realised and unrealised.

## Balance Sheets

<i>Assets</i>		<i>Liabilities and Net Worth</i>	
OPENING BALANCE SHEET			
Nonfinancial assets.....	9,922	Liabilities	6,298
Financial assets.....	6,792	Net worth.....	10,416
CLOSING BALANCE SHEET			
Nonfinancial assets.....	10,404	Liabilities	6,955
Financial assets.....	7,522	Net worth.....	10,951

The *Opening* and the *Closing balance sheets* complete the sequence of accounts, rendering the *SNA 1993* stock-flow-consistent. Net worth is equal to the difference between total assets and liabilities at each point in time; and the change in net worth over time ( $10,951 - 10,416 = 535$ ) can be calculated from the accumulation accounts ( $535 = 230 + 17 + 288$ ).

Two additional accounts have to be added to close the system. First, the *Rest of the world account* records all flows between residents and non-residents as well as the related stocks. The 'rest of the world' is simply treated as an additional sector; and the sequence of accounts discussed so far is advanced for this sector, too, albeit in a simplified form. Only two current accounts are set up. The *External account of goods and services* records imports, exports, and, as the balancing item, the external balance. If a nation's external balance is negative (its exports are lower than its imports), this negative

balance is a *positive* resource (addition to value) for the second current external account, the *External account of primary incomes and current transfers*, since the recording for the *Rest of the world account* is from the point of view of that pseudo-sector. After all other relevant cross-border current-account flows have been added and deduced, the balancing item is the 'current external balance'. The external accumulation accounts mainly record cross-border financial flows; and the final account shows the financial assets and liabilities held by non-residents.

Whereas the External account of primary incomes and current transfers offsets records in the Allocation of primary income account (see above), the entries in the External account of goods and services have still to be offset. The same holds good for the domestic transactions in goods and services. This omission is corrected in the final account of the sequence, the *Goods and services account*.

## Goods and Services Account

<i>Resources</i>		<i>Uses</i>	
Output.....	3,604	Intermediate consumption.....	1,883
Taxes less subsidies on products....	133	Final consumption expenditure.....	1,399
Imports of goods and services.....	499	Gross capital formation.....	414
		Exports of goods and services.....	540
Total.....	4,236	Total.....	4,236

The *Goods and services account* records entries reversed from the usual sides: uses are right while resources are left. Thus, every entry in the whole system has found its balancing counterpart. *The SNA 1993 constitutes a fully integrated system of economic accounts.* This integrated system of accounts forms the core of the *SNA 1993*.

#### *Satellite Accounts*

The downside of the *SNA* being a fully integrated system of economic accounts is that some information, although of social concern, cannot be displayed in the system because it does not lend itself to double-entry bookkeeping. Generally speaking, the *SNA* framework is less appropriate where no market for a good or service exists. In fact, some non-market transactions that have found their way to the *SNA* core, e.g. public services, although they are beset with problems of valuation. Still, there are other non-marketable goods and services providing utility to the public which should be recorded somehow. This kind of information is relegated to so-called satellite accounts that form a part of *SNA*, but not of its fully integrated core. The most important objective of satellite accounts is to broaden the perspective in order to gain a fuller picture of the 'welfare' of a nation than the *SNA* core with its focus on (paid) *economic* activities will convey. In many countries, satellite accounts are constructed for subjects such as household production (e.g. child care, cooking, cleaning etc.), research & development, public health, and the quality of the environment (Eisner 1996).

#### *Supply and Use and Input-Output Tables*

The *SNA*'s central framework presents data on a highly aggregated level. For many

analyses, however, it would be desirable to have these data disaggregated by industry, and by type of product. This disaggregation is done in the 'supply and use tables', which, again, form part of the *SNA*, but not of its core. The supply and use tables present a *Production account* and a *Generation of income account* by industry as well as the *Goods and services account* by product. Supply and use tables are usually constructed as matrices with industries as columns and products as rows. These matrices can be transformed in such a way that it becomes possible to read off all inputs every industry needs to produce its specific output, hence the name 'input-output tables' (cf. *SNA 1993*, Ch. 15).

### **National Accounts and Governance**

#### *Usefulness of National Accounts for Governance*

"The chief impetus to the development of economic accounts has come from central governments, which probably remain their chief users. By monitoring economic movements, policy-making agencies including the central bank can see if they are on track with respect to national objectives regarding growth, price inflation, the trade balance, unemployment, and so on, and, if not, they can take appropriate actions" (Kendrick 1996:4-5).

National accounts are the main source of information about the state of the economy. Their data serve as input for growth predictions and business cycle forecasts, which are usually made with the help of intricate econometric models and techniques. Also, medium-term budgeting is typically done within the framework of national accounts. It has to be stressed, though, that national accounts synthesise

data usually collected for other purposes (e.g. tax collection). A lot of estimation is involved in calculating the aggregates entering into the *SNA*. Therefore, data revisions are common and sometimes substantial—and national accounts may have a time lag of up to two years for final estimates. So, unlike meteorologists, applied economists normally do not know the present day situation and not even the recent past when making predictions.

This might be one reason why business cycle forecasts based on national accounts data are generally regarded as being unsuccessful in a medium-term horizon (of more than one year or so). Also, *Kendrick's* optimism as to the possibility that “the economic relationships revealed by analyses based on the income and product accounts make it possible to predict the consequences of given or alternative policy actions” (1968:33) is now widely discarded, following the influential ‘Lucas Critique’ (*Lucas* 1976). Nevertheless, national accounts have still a significant impact on economic governance. Growth, productivity, and inflation forecasts serve as input into the wage bargaining process and thus affect the disposable income of households.

The so-called ‘functional’ distribution of income between labour and capital can be read from the *Generation of income account*, though it is not possible to gather information about the ‘personal’ distribution of income (between different types of households) from the *SNA*. Price data are also crucial for the alignment of retirement pensions in countries with pay-as-you-go-systems, where pensions are automatically increased in line with inflation. (The same holds good for all price-indexed contracts such as – in some countries – mortgage loans.) Elsewhere,

e.g. in Germany, retirement pensions should rise to the same extent as after-tax wages, so here we need wage data to calculate pension alignments. Both price and wage data could, in principle, be gathered from national accounts. In practice, though, the time-lag in publication is too long, so that wage and price data are taken from other statistical sources such as consumer price and wage indices. These then serve as input for the construction of the ‘real’—or constant price—time series in national accounts.

Price data and inflation forecasts are also indispensable for gearing monetary policy. It is now consensus that the main task of central banks is to keep inflation under control. Since monetary policy has a time lag of up to three years in some countries, the central bank needs to forecast what will be the rate of inflation in the future if it keeps the interest rate unchanged. Central banks use regression equations and other models for that purpose. National accounts data not only enter into the coefficients of these models. Also, projections of national accounts data, e.g. of demand variables, are necessary to produce a forecast for the rate of inflation. If such a forecast signals that inflationary pressures will emerge in three years time, then, if this is the length of the monetary policy lag, the central bank will raise the rate of interest now. So, projections for national accounts data have an impact on current interest rates and, hence, rentier income. Also, the exchange rates of the domestic currency will be affected. An interest rate hike leads to an appreciation of the currency in the short run.

This has an adverse effect on the balance of trade, unless the development of the terms of trade acts as a counterbalance. The terms of trade are a measure for the

overall competitiveness of the domestic economy. They are calculated as the ratio of export prices over import prices (which can both be read from the *Goods and services account*). If the terms of trade improve—export prices rise faster than import prices—then this can be interpreted as a substitute for domestic growth. The domestic economy may or may not have grown, but at any rate a greater basket of goods (and services) can be bought from abroad in exchange for the output of the domestic economy. Hence, consumption opportunities at home have improved. On the other hand, if export prices rise faster than import prices, this could impair the competitiveness of domestic exporters on the world market. In a small open economy, such a development could prompt the central bank to cut interest rates – aiming at a devaluation of the domestic currency to support the export industries.

If, in a country, fiscal policy follows an activist approach, then it will react to an unsatisfactory growth or business cycle outlook by taking discretionary measures. Traditional ‘Keynesian’ measures, i.e. deficit spending, are now widely out of fashion, especially in Europe, because it is believed that they irresponsibly add to public debt and thus overburden future generations. What is more, they are beset with the ‘time inconsistency problem’, that is, it takes so much time to decide on them (in the parliamentary process) that the phase of the business cycle will have changed until they become effective. So, their repercussions will be pro-cyclical rather than anti-cyclical. As a consequence of this view, fiscal policy now restricts itself to improving the general set-up of the economy (the supply conditions) in many countries in order to tackle dim growth prospects. This may involve tax-cutting or

deregulation policies that aim at fostering competition and thus the competitiveness of the domestic economy.

The mistrust of discretionary fiscal policy has become so great that many countries have decided to adopt rules of conduct for their government politicians. One example is provided by the ‘Maastricht criteria’ which prescribe, amongst other things, that in (potential) member countries of the ‘European Monetary Union’ (EMU) the rate of consumer price inflation shall not exceed 2 percent, that the ratio of new public indebtedness to nominal GDP shall not be higher than 3 percent, and that the ratio of gross public debt to nominal GDP shall not be higher than 60 percent. These criteria which, in practice, place considerable constraints on both monetary and fiscal policy in the EMU, are all checked using (early estimates of) national accounts data. Switzerland, to give another example, has adopted an expenditure rule that necessitates the planned expenditures of the federal government to be equal to the projected receipts multiplied by a ‘business cycle factor’. This factor is calculated as the ratio of potential output over next year’s forecasted output. Again, public expenditures depend on projections of national accounts data.

Private businesses can also benefit from national accounts. Input-output matrices “can be converted to employment and capital requirements ... The projections are useful not only to enterprises in the various industries but also to governments in planning outlays for infrastructure, education, training and retraining, and so on” (Kendrick 1996:4). Even the demand for different types of financial instruments can be estimated from the *Financial accounts* data.



### *Abuse of National Accounts: Governance*

Under this heading, we would discuss the use of national accounts for international comparisons. This must be surprising—since national accounts have been conceived right from the beginning to serve this very purpose. What should we need an internationally agreed-upon standard – the *SNA* 1993—for if not to allow for international comparisons? Indeed, the main impetus for the first *SNA* came from the lack of an objective criterion on which to base the dues of the United Nations. These are now calculated by a formula that includes per capita income—a national accounts magnitude. The same holds good for the contributions that its member states have to transfer to the European Union. Also, the World Bank assesses eligibility for loans or aid on the basis of a country's per capita income.

The balance of the *Allocation of Primary Income Account*, i.e. gross national income (GNI), serves as basis for comparisons of rich countries' development aid payments. In 1970, the United Nations recommended that rich countries should spend 0.7% of their GNI for development aid, but, in fact, except for some Scandinavian and Low countries the developed world does not abide by that rule. – We here do not deny that international comparisons are often suggestive and sometimes inescapable. (Data sources include Heston et al 2002; Maddison 2003.) Yet, the way in which international comparisons on the basis of national accounts data are often carried out amounts to an abuse of this framework that might lead to questionable policy conclusions.

Let us take the debate about economic growth as an example. As a matter of fact,

many OECD countries are experiencing declining growth rates of real GDP. While, from a theoretical point of view, international convergence in per capita income should be expected—so that declining growth rates in rich countries are to a certain extent normal (cf. Mankiw et al 1992)—there is one radiant counter-example: the United States. Over the course of the 90s and up through the new millennium, the U.S. have outperformed all other high-income countries with respect to growth. National accounts data show this clearly. So we must conclude that other rich countries could also grow faster if only their institutions were adequately designed. At least, this conclusion has been drawn in many European countries where public debate centres on the question how to devise 'structural reforms' in order to make the set-up of the respective economy more similar to that of the United States. Tax-cutting, reducing the size of the public sector, and 'labour market reforms' (e.g. introducing a low-wage sector, curtailing the influence of trade unions, and increasing either weekly or lifetime working hours) are typical ingredients of this kind of debate, which is also promoted by international organisations such as the OECD, the IMF, and the WTO. The main impact of national accounts on governance can be found here. And this impact is strong; it has already changed many societies—and even more changes can be expected for the future.

We shall not discuss the theoretical validity of the arguments behind the debate just sketched out. Of course, the better growth performance of the U.S. could have other reasons than a more flexible labour market etc. Our point here will be that a careful analysis of the relevant national accounts data reveals part of the headstart

of the United States in growth as a statistical artefact.

Let us compare U.S. per capita growth with the OECD's slowest-growing economy, Switzerland. See Figure 2, below:

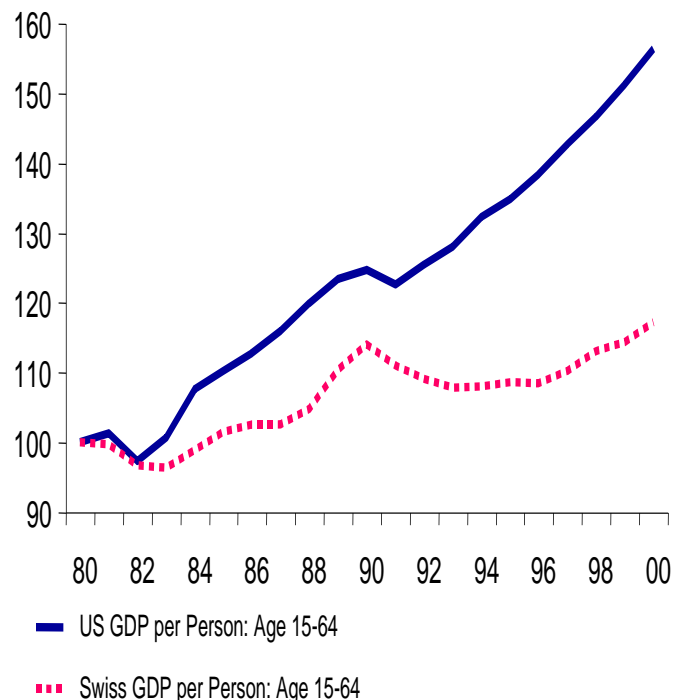


Figure 2: U.S. and Swiss per capita GDP at constant prices, 1980 = 100 (Source: OECD, back issues)

For international comparisons, time series at constant prices are preferred because they show ‘real’ growth, whereas the series at current prices might have been pushed up by inflation. But this prompts the question how the deflators are constructed. Unfortunately, there is no agreed-upon international standard for the construction of deflators. Countries that lay greater emphasis on alleged quality improvements will interpret an observed rise in prices to a lesser extent as ‘inflationary’ and to a greater extent as a reward for improved quality. Hence, the deflator will rise to a lesser, and the ‘real output’ to a greater extent than in other countries. Now, in the U.S., deflators are

‘quality adjusted’ in this way using the so-called ‘hedonic method’, whereas in Europe, including Switzerland, they are not. A significant portion of the U.S. headstart vis-à-vis Europe with respect to real growth seems to be attributable to this difference. Also, note that there is more justification for the central bank to keep the interest rate down if the rate of inflation—as published by the statistical office—is low. Since a low interest rate supports domestic demand, the choice of the statistical concept for calculating deflators can have real consequences for growth.

Yet, there are additional problems in comparing the two GDP figures. First, the data for Switzerland are according to *SNA 1968* because the Swiss Federal Statistical Office has only recently accomplished the change to *SNA 1993* and, at the time this entry was written, new GDP figures back to 1980 were not yet available. The data for the U.S., on the other hand, are according to *SNA 1993*. As was shown in Figure 1, U.S. GDP according to *SNA 1968* was lower. Still, for a meaningful international comparison, all data have to conform to the same system. Figure 3 adds two time series to Figure 2. U.S. GDP is now also calculated according to *SNA 1968*. (This time series ended in 1997.) Two additional adjustments are made. First, GDP is now related to hours worked instead of persons. In the U.S., the number of hours worked per year and employee increased slightly from 1980 to 2000, while in Switzerland it fell by as much as 13%.

Secondly, growth in the terms of trade is also taken into consideration. As was explained earlier, even in a country where real value added does not increase over time, an increase in overall wealth can be observed if, due to trade with foreign countries, more import goods can be

acquired by paying with the same amount of domestic products. Since 1980, the terms of trade have improved both in the U.S. and in Switzerland, but much more so in Switzerland. When combining the effects of the SNA revision, of the different

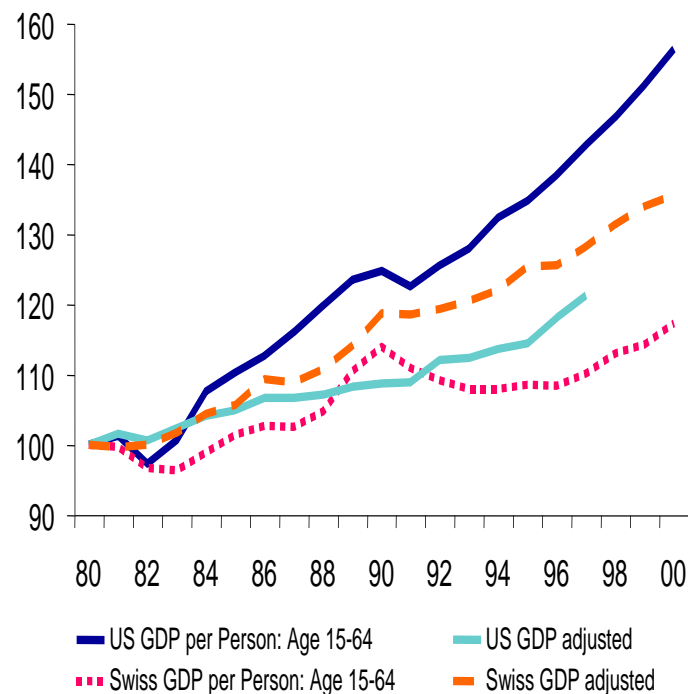


Figure 3: Terms of Trade and SNA adjusted U.S. and Swiss GDP per working hour, 1980 = 100 (Sources: OECD, back issues, Heston et al., 2002, University of Groningen and The Conference Board, 2002)

While we would admit that the latter contrast is perhaps as misleading as the former, this example highlights the limitations of national accounts for purposes of international comparisons. Admittedly, the U.S. have grown more, but the citizens also had to increase their efforts (measured in labour time) to realise this growth. The foregone utility of leisure, though, is nowhere recorded in the System of National Accounts. This evidences again what has been pointed out above – that the SNA core with its focus on paid economic activities does not give a full picture of the welfare of a nation. Hence, proposals for institutional reforms that come along with

dynamics of labour inputs, and of the terms of trade changes, then growth over the period between 1980 and 1997 was actually higher in the slowest-growing economy—Switzerland—than in the fast-growing U.S. (cf. Abrahamsen et al 2003). multi-country comparisons of growth rates should be handled with care.

### Selected References

- Abrahamsen, Yngve, Roland Aeppli, Erdal Atukeren, Michael Graff, Christian Müller, and Bernd Schips. (2003) “The Swiss Disease: Facts and Atefacts”, KOF Working Paper, No. 71.
- Brümmerhoff, Dieter. (2002) *Volkswirtschaftliche Gesamtrechnungen*, Seventh Edition. Oldenbourg Verlag, München.
- Carson, Carol S. (1996) “Design of Economic Accounts and the 1993 *System of National Accounts*”, in John W. Kendrick (Editor), *The New System of National Accounts*. Boston: Kluwer Academic Publishers, pp.25-72.
- Eisner, Robert. (1996) “Expansion of boundaries and Satellite Accounts”, in John W. Kendrick (Editor), *New System of National Accounts*. Kluwer Academic Publishers, Boston, pp.91-113.
- Hartwig, Jochen. (2001) “Petty—oder: die Geburt der Arbeitswertlehre aus ökonomischen Problemen des frühen Kapitalismus”, *Historical Social Research/ Historische Sozialforschung*, Volume 26, Number 4, pp. 88-124.
- Heston, Alan, Robert Summers and Bettina Aten. (2002) *Penn World Table, Version 6.1*. Center for International Comparisons at the University of Pennsylvania (CICUP).
- Inter-Secretariat Working Group on National Accounts. (1993) *System of National Accounts 1993*.

- Brussels/Luxembourg: Commission of the European Communities/Eurostat, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, World Bank. (cit.: *SNA 1993*).
- Jackson, Dudley. (2000) *The New National Accounts—An Introduction to the System of National Accounts 1993 and the European System of Accounts 1995*. Cheltenham, UK: Edward Elgar.
- Kendrick, John W. (1968) “National Income and Product Accounts”, in David L. Sills (Editor), *International Encyclopedia of the Social Sciences*. Volume 11. London: Collier-Macmillan, pp. 19-34.
- Kendrick, John W. (1970) “The Historical Development of National Income Accounts”, *History of Political Economy*, Volume 2, Number 3, pp. 284-315.
- Kendrick, John W. (1996) “Introduction and Overview”, in: John W. Kendrick (Editor), *The New System of National Accounts*. Boston: Kluwer Academic Publishers, pp. 1-23.
- Kuznets, Simon. (1934) *National Income 1929-1932*. Washington DC: US Congress, S. Doc. 124, 73<sup>rd</sup> Congress, 2<sup>nd</sup> Session.
- Lucas, Jr., Robert E. (1976) “Econometric Policy Evaluation: A Critique”, in: Karl Brunner and Alan H. Meltzer (Editors), *The Phillips Curve and Labor Markets, Carnegie-Rochester Conference Series on Public Policy, Volume 1, Journal of Monetary Economics: Supplementary Issue*, pp. 19-46.
- Maddison, Angus. (2003) *The World Economy: Historical Statistics*. Paris: OECD.
- Mankiw, N. Gregory, Romer, David, and Weil, David, N. (1992) “A Contribution to the Empirics of Economic Growth”, *Quarterly Journal of Economics*, Volume 107, Number 2, pp. 407-437.
- OECD. (2007) *National Accounts of OECD Countries*. Paris: OECD.
- Stone, Richard. (1947) “Functions and criteria of a system of Social Accounting”, *Income and Wealth*, Series I, pp. 1-74.
- Stone, Richard. (1951) “Definitions and Measurement of the National Income and Related Tools”, in: *Measurement of National Income and the Construction of Social Accounts*, Studies and Reports on Statistical Methods. No. 7. Geneva: United Nations.
- Studenski, Paul. (1958, 1961) *The Income of Nations*, 2 Volumes. New York: New York University Press.
- United Nations. (1953) *A System of National Accounts and Supporting Tables*. Studies in Methods, Series F, No. 2. New York: United Nations.
- United Nations. (1968) *A System of National Accounts*. Studies in Methods, Series F, No. 2, Rev. 3. New York: United Nations.
- University of Groningen and The Conference Board. (2002) *GGDC Total Economy Database*. <http://www.ggdc.net>
- Yngve Abrahamsen and Jochen Hartwig, *Swiss Institute for Business Cycle Research, Zürich*.

## Pareto Optimality, Second Best and Social Welfare

Michael McLure

### Introduction

The proposition that society experiences an improvement in welfare when a change in the social state provides a benefit to at least one individual, without adversely affecting any members of society, is usually referred to as the ‘Pareto principle’. This principle was initially introduced to investigate the impact of market activity on collective ‘economic welfare’, culminating in the *first law of welfare economics*, namely the proposition that an equilibrium state attained under free competition is Pareto optimal. This is justified on the basis that any movement from that state will necessarily violate the Pareto principle. The Pareto principle has subsequently been employed in a wider context and is now utilized as a benchmark for the development of public policy options and for evaluating the impact of government decisions on collective ‘social welfare’.

The representation of benefit or harm experienced by individuals under different economic or social states is fundamental to the theory of collective welfare. In principle, benefit or harm depends on whether an individual subjectively determines that a change in their utility, from movement between different economic or social states, is positive or negative. As the Pareto principle depends on the sign of a change in utility, and not the actual magnitude of the change in utility, the *New Welfare Economics* inspired by Hicks (1939) came to formulate welfare theory in reference to individual’s ordering of preferences over different states. By and large, this ‘ordinal’

approach has replaced the old welfare analysis where utility is specified as a quantifiable magnitude. In the case of economic welfare, positive or negative variations in utility are usually associated with variations in the pleasure that an individual experiences from the satisfaction of tastes. In the case of social welfare, it is explicitly recognized that individuals have ethics and preferences that are not necessarily ego centric or hedonistic. As a consequence, individuals may experience a personal subjective gain or loss from the impact that different social states have on all members of the collective, not simply themselves.

### Collective Economic Welfare

The first law of welfare economics has its origins in Francis Ysidro Edgeworth’s *Mathematical Psychics* (1881). On the utilitarian presumption that subjective utility is measurable by each individual, Edgeworth introduced a theory of exchange that represents equilibrium under free competition as a relative economic maximum. However, individuals’ subjective quantifications of utility are not interpersonally comparable, at least not in any objective and consistent sense, and it was Vilfredo Pareto (1894) who introduced production to the analysis of welfare to demonstrate, without relying on interpersonal comparisons of utility, that equilibrium under free competition represents an economic maximum. On observing that economic goods are positively, but un-quantifiably, correlated with each individual’s assessment of utility, Pareto suggested that any arbitrary commodity, a ‘numeraire’, could provide the basis for representing an individual’s ordinal index of utility: the number of numeraire goods with the same purchasing



power as an individual's bundle of consumer goods and productive services represents provides an ordinal index of the individual's utility.

An individual experiences a gain (loss) when the numeraire equivalent of the difference between total utility gained from consumption services and total utility spent on the consumption of productive services rises (falls). When productive techniques approach those that prevail under free competition, productivity increases and, initially, everyone may experience an increase in utility because purchasing power rises and utility spent on productive services is largely unchanged. In these circumstances, there is no need to resort to interpersonal comparisons of utility as the unequivocal increase in everyone's net purchasing power rises attests to an improvement of collective economic welfare. However, before attaining a freely competitive equilibrium, Pareto showed that it is almost certain that some members of the collective will be harmed when others realize a gain. Importantly, even though such change violates the Pareto principle, he did not consider this an economic maximum because further change towards the cost minimizing production techniques would provide the collective with the potential to fully compensate those who were harmed by such changes. Only once production techniques accord with those that would prevail under free competition, where production costs are minimized, has an economic maximum been achieved, as any change from that point would necessarily violate the Pareto principle without offering the potential for possible losers to be fully compensated by the collective for their loss.

Pareto (1896/97, 1909) subsequently commenced analysis of collective economic welfare from a point of equilibrium established under free competition, enabling what is now known as the first law of welfare economics to be expressed without resort to the 'compensation principle'. Seminal studies associated with the identification of the necessary conditions and proofs of the first law of welfare economics include Abba Lerner (1932), Oskar Lange (1942) and Maurice Allais (1943), and Kenneth Arrow (1951).

The modern demonstration of Pareto optimality generally centres on three necessary conditions: (i) efficient exchange; (ii) efficient production; and (iii) efficient composition of output. Failure to achieve all three of these conditions prevents the realization of a Pareto optimal allocation. Efficient exchange requires the maximization of each consumer's utility from every exchange transaction, which is achieved when the relative price of goods equates to their relative marginal utility (or equality of marginal rate of substitution in *New Welfare Economics*). Efficiency in production requires the minimization of utility spent from the consumption of productive services, which is achieved when the relative price for productive services reflects the relative marginal utility spent on the consumption of productive services during production (or equality of the marginal rates of technical substitution between productive services in *New Welfare Economics*). Finally, efficiency in the composition of economic output ensures the maximization of profits from production by requiring that the composition of output reflect the demands of the market, subject to efficiency in exchange and production. This condition is

achieved when the relative price of goods equals their relative marginal costs (or the marginal rates of substitution equal the marginal rate at which productive services are transformed into goods in *New Welfare Economics*). Consequently, if the price of a good (X) is twice that of another good (Y) and this is Pareto optimal: the marginal benefit from consuming good X is twice that of good Y (exchange efficiency); the cost of producing good X is twice the marginal cost of producing good Y (efficiency in the composition of output); and the value of productive services consumed in the production of goods X and Y has been minimized (production efficiency). If these conditions are not met, then a Pareto improvement in economic welfare is possible.

More recently, the concepts of efficiency in exchange, production and composition of output, which collectively define Pareto optimality, have been complemented by the notion of 'dynamic efficiency'. As Pareto optimality is a timeless notion, it does not assign any role to innovation and productivity enhancements that may emerge from the process of competition. Dynamic efficiency, an aspect of the broader notion of 'endogenous growth', has come to represent an important element in the case for increasing global competition, on the basis that the research and innovation resulting from the competitive process will increase global economic welfare.

### **Fairness and Economic Welfare**

The first law of welfare economics is not unconditional, as it concerns the realisation of a Pareto optimal allocation subject to the initial distribution of resources. As Pareto (1894) stressed, the proposition that free competition is optimal "concerns

production and not the distribution of economic goods". Theorists who first considered the question of optimality generally acknowledged this limitation and, concurrent with the development of the first law of welfare economics, a second law of welfare economics emerged to address the relationship between different resource endowments and the notion of Pareto optimality.

If an initial endowment were to change, free competition would yield a new equilibrium. The first law suggests that this too would also be Pareto optimal. Alternative initial distributions create alternative points of Pareto optimality as variations in the initial distribution of resources alter the pattern of demand for consumer and producer goods. By working backwards from this point, the second law of welfare economics established that any possible Pareto-optimal allocation can be achieved through competitive equilibrium given the appropriate initial allocation of resources, or given the appropriate implied initial distribution of resources after lump sum transfers have been imposed on firms and individuals. The importance of "lump sum" transfers is that they effectively alter an initial distribution of resources without having any direct impact on relative prices. As a consequence, the second law of welfare economics is policy relevant—it establishes that governments can utilise their fiscal system to, in effect, alter the initial distribution of resource without altering relative prices, and that free competition under the new resource allocation will be Pareto optimal.

The second law of welfare economics was directly linked to research into the economics of socialism, which commenced with Pareto's insistence that, from pure economics alone, the superiority of

capitalism or socialism cannot be established because both are, in principle, capable of achieving a relative economic maximum. In the *Cours d'Économie Politique* (1896-97), his analysis suggests that collective economic welfare could be maximized under socialism when: (i) the 'Minister for Production' approves production techniques based on the functions of input prices that would apply under free competition; and (ii) the 'Minister for Justice' determines the distribution of product in accord with the government's view of an equitable outcome. However, Pareto's interpretation of his own analysis was partly flawed (Chipman 1976). It was Enrico Barone (1908), and subsequently Oskar Lange (1938, 1942), who more fully developed the economics of socialism, although their approaches have been criticised for making socialism function as if it were capitalism (King 2003) and for ignoring the difficulty (even impossibility) of achieving the level of planned and coordinated production necessary for Pareto optimality to be achieved (Hayek 1938).

Nevertheless, the second law of welfare economics remains important for policy purposes as it suggests that governments are able to alter distribution of resources and still achieve a Pareto optimal outcome. From a policy perspective, perhaps the most important aspect concerns the application of the 'compensation principle' to establish whether government decisions to alter society's distribution and level of available production possibilities enhance or diminish collective economic welfare. A change in production possibilities of a society ensures that the initial resource endowment in the current state (A) and the alternate state (B) are different, and the composition of goods demanded in each

state will vary accordingly. As any movement between the two states will necessarily violate the Pareto principle, the compensation principle is necessary to facilitate welfare comparisons. However, this only allows such comparisons in very limited circumstances.

Tibor Scitovsky (1941) pointed out that, as a consequence of the different composition of demand associated with the two different initial resource endowments, one cannot rule out welfare reversals, where movement from state (A) to state (B) provides the potential for those who gain to fully compensate losers, while at the same time, a return to state (A) from state (B) may also provide the potential for those who gain to fully compensate losers. Therefore, comparisons of states subject to welfare reversals cannot be undertaken using the compensation principle. Furthermore, the compensation principle cannot be utilised in the face of intransitivities in community welfare orderings, such as when a community prefers  $x$  to  $y$  and  $y$  to  $z$  under state (A) and prefers  $z$  to  $x$  under state (B). As intransitivities are, like welfare reversals, feasible when comparing community preferences across economic equilibria founded on different initial endowments, the legitimate use of the compensation principle is constrained significantly.

In light of this, choices pertaining to distribution cannot be effectively considered in isolation from social issues that do not feature in (or are masked and muted in) market behaviour. Consideration of collective economic welfare must be supplemented by a broader examination of collective social welfare to achieve an efficient and fair outcome. Government policy is influential in shaping 'social welfare' as it is a practical instrumental of



social choice. However, before progressing to the question of collective 'social welfare', it is appropriate to note some limitations of the laws of welfare economics associated with the endogeneity of preferences.

### **Endogenous Preferences**

The first law of welfare economics is predicated on the view that individuals' utility functions are conserved over time, even as the economic and social state changes. On this basis, utility is presented as independent of the path that a particular society follows in arriving at a certain state. The second law is simply a logical consequence of the first law. As such, it too is predicated on 'path-independent' utility. However, just as preferences motivate human action, preferences may also alter in response to human action. This endogeneity of preferences ensures that utility is not strictly path-independent, so history, institutions and time matter. This poses issues for public governance and public policy. There are three main types of response to the recognition of preference endogeneity.

First, challenge the laws of welfare economics in light of endogenous preferences. This approach is typified by *A Quiet Revolution in Welfare Economics* (Albert and Hahnel 1990), which provides a critique of welfare economics and lays the foundations for re-orienting the field of study by recognising the developmental benefits of human centred activity and accounting for institutional boundaries that interact with, and constrain, such activity. However, it must be stressed that work along this line is not yet well developed.

Second, adopt a purely individualistic theoretical specification of economic and political conduct. In this regard, the

constitutional approach to political economy associated with James Buchanan (1962, 1987) considers the second law of welfare economics to be of very limited relevance because it requires collective accord on the appropriate distribution of resources. Instead of relying on the laws of welfare economics to inspire public policy, constitutional political economists apply the Pareto principle directly to the constitutional rules that govern the decision-making framework for economic policy. Agreement with such rules that govern political conduct by every individual member of the collective is considered optimal, even when subsequent policy decisions made in accordance with these rules benefits some while harming others. The policy relevance of this approach is that liberal governance becomes the central factor in determining whether government's contribution to the economy is efficient and fair.

Third, accept the fundamental laws in the context of static economics, although recognise that supplementary investigations into dynamic efficiency are also necessary to provide a basis for competition policy. Within this context, the first law is seen as most relevant to policy issues when correcting failures in specific markets. However, to establish the need and extent of such policy action, it is necessary to determine whether a first best or second best solution is appropriate.

### **First Best or Second-Best?**

Two broad economic welfare approaches to governments' role in maximizing collective economic welfare have emerged: the first best and the second best approaches. First best economic policy is predicated on the view that collective economic welfare is enhanced by

government action that brings general markets arrangements closer to what would prevail if markets were competitive. As each incremental movement in that direction is expected to improve collective economic welfare, first best policy prescriptions generally attempt to make markets work more competitively to enhance efficiency. In contrast, second best economic policy does not automatically accept that an increment in the competitiveness of markets is efficiency enhancing (unless such an increment realises perfect competition). Under this approach, some market imperfections are considered unavoidable and perfect competition across all markets is unattainable. Once these unavoidable market imperfections are recognised as a constraint, and economic interdependencies related to these constraints are examined, maximizing collective economic welfare may require policies that do not lead to an increment in competition, governments may introduce second best policies to maximize economic welfare in light of constraints caused by unavoidable market imperfections.

A government implementing first best policies, based on the fundamental theorems of welfare economics, may intervene in private economic activity when: markets are imperfectly competitive; competition fails to achieve efficient outcomes; or when it is efficient to redistribute resources. When a market has a limited number of suppliers and there is no evidence that economic forces from 'potential' competitors will impose a discipline on suppliers to act competitively, then a first best policy may also regulate the market in an attempt to make it function more competitively. When competitive markets do not achieve

efficient outcomes, the failure is often due to two related factors. Either the market fails to account for: net external costs or benefits that accrue to parties not directly involved in a particular market process; or for the "social" attributes of some goods which provide for collective consumption.

Externalities were recognised as a source of market failure in Sidgwick's (1883) *Principles of Political Economy* (see Groenewegen 2003), but the analytics of the relationship between markets and externalities was not addressed systematically until Arthur Cecil Pigou's seminal *Economics of Welfare* (1920). Pigou's analysis suggested that an efficient outcome for a particular market requires external costs and benefits, transferred to other markets or to non-market assets like the environment, to be internalised by the market and reflected in relative prices. Traditionally, public policy has been considered the appropriate means of achieving this goal, with coercive pigouvian taxes imposed on economic agents who generate externalities that impose costs on the community, and pigouvian grants made to economic agents who generate externalities that impart net benefits on third party members of the community, introduced to ensure that price signals in particular markets reflect full economic costs. For example, a producer (the first party) may manufacture a product for sale to consumers (the second party) using a method which damages the environment used by other members of the community (the third parties). As a result of this 'market failure', the loss experienced by third parties exceeds to gain by the first two parties. A coercive Pigouvian tax imposed on the first party would directly remove the producer's gains and indirectly remove the gain experienced

by the second party, as producers increase prices to consumers in response to the tax. The funds collected from the tax would then be available to compensate third parties for their loss from the pollution of their environment. An alternative solution, associated with the work of Ronald Coase, is for property rights to be fully assigned to broaden the market and internalise externalities within the market structures themselves, which will be Pareto optimal as long as there are no new transaction costs introduced as a result of the re-assignment of property rights. In the above case, the third party would have the right to demand compensation directly from the first party, thereby removing the need for Government to correct the market failure.

The policy response to externalities is further complicated when economic activity affects parties in more than one country. In such circumstances, the net economic cost or benefit from a particular externality will generally be different across nations and the possibility of Pigouvian solutions is reduced because coordination is required between different national governments. The potential for Coasian type solutions will also be reduced unless governments don't provide for property rights to protect injured parties from damage caused by the actions of parties in other countries. Furthermore, global externalities are not always the makings of private parties. War and financial stability may generate global externalities that substantially derive directly from the actions of national governments.

Goods are 'public', or social, when their consumption is non-rival, in that an individual's consumption has no impact on the benefit that others receive from contemporaneous consumption of the same

good, and non-excludable, in that actions by one individual cannot prevent other individuals from consuming the good. Such goods may be 'public' at local, regional, national or global levels. The inability of private markets to efficiently provide public goods was first explicitly recognised in the late 19<sup>th</sup> Century by Emil Sax (Austria) and Ugo Mazzola (Italy), with public provision seen as potentially efficient. The precise formulation for efficient public good provision was formally developed by Paul Samuelson (1954,1955), who demonstrated that Pareto optimal provision of public goods is achieved when the sum of non-rival benefits that all individual experience from the consumption of an additional unit of a public good equals the cost of supplying that marginal unit. This suggests that efficient public good provision requires individuals to contribute, via the tax system, an amount to fund the provision of public goods that is equal to the benefit they receive from consuming public goods. The most difficult policy issue concerning the provision of public goods is the under-identification of individuals' preference for public goods as, contrary to the case of private goods, individuals may choose to not reveal their true valuation of a public good as a strategy for receiving a benefit that is paid for by others members of the collective.

As with the issue of externalities, the spatial dimension to public goods presents particular public policy problems. When the non-rival and non-excludable aspects of a publicly provided good are localised, then the fiscal competition between local governments can enhance efficiency if the local population is mobile enough for rate and tax payers to vote with their feet (Tiebout 1956). However, when the

‘publicness’ (or the non-rival and non-excludable attributes) of a good extends nationally, efficiency enhancing public policy responses shift, away from fiscal competition, towards coordination by a single public authority acting in response to the preferences of individual members of the collective. When the ‘publicness’ of a good (or bad, such as the case of greenhouse gases) is global, efficient policies rely on co-ordination between national governments

In contrast to the above focus on competition, together with coercion, co-ordination and co-operation to correct for market failure and the publicness of goods, the second best approach, initiated by Lipsey and Lancaster (1956), focuses on the conditions of Pareto optimality. If there is efficiency in exchange, production and output composition, then a first best outcome can be realised. However, when any one of these three components of efficiency is not met, then there is no presumption that meeting the other two components will enhance efficiency. Moreover, this proposition is not restricted to one sector of the economy; it extends across all sectors as the economy is interdependent. This, in part, adds weight to the proposition that some degree of inefficiency is inevitable, because markets are not perfect, and that second best policies should be the norm rather than the exception. Yew-Kwang Ng (1990) has argued that the policy relevance of second best approaches depends on the relative abundance of ‘information’ available on welfare relationships under first and second best approaches. When such information is unavailable or scarce, the cost of acquiring information necessary for second best policies increases and the chance of improving net welfare through first best

policies increases. Alternatively, when such information is available, Ng advocates a ‘third best’ approach, or effectively a second best analysis that accounts for all the costs of acquiring information necessary to develop second best policy. However, John Cullis and Philip Jones (1992) report that Ng has been criticised for treating information costs as an input cost, rather than as an additional constraint on production, and because the market imperfections used to justify second best policy may have been caused by public policy at an earlier time.

### **Collective Social Welfare**

One fundamental factor to emerge from general consideration of collective economic welfare is that, in the end, synthetic judgments of an ethical nature concerning the welfare of the collective are inevitable, either implicitly or explicitly. If the first law of welfare economics is accepted as the dominant policy criterion, the prevailing distribution of resources and social relations are implicitly endorsed ahead of all alternative social states. Alternatively, if both the first and second welfare laws are accepted as the dominant policy relevant criteria, the Pareto principle alone does not permit a complete comparison of collective welfare in different states. While the compensation principle permits clear and consistent welfare findings when comparing Pareto optimal allocations against non-Pareto optimal allocations under different initial resources allocations (or different initial resource allocations modified through lump sum taxes and grants), it is incapable of facilitating complete comparisons of collective economic welfare for different points of Pareto optimality, as changes in the initial resource allocation lead to

welfare reversals and intransitivities in community preferences.

Consequently, even in a first best world, some judgment is required to determine which Pareto optimal allocation, from the range of possible relative economic maxima, will maximize collective social welfare. This was first clearly recognized by Pareto (1913), who provided the “first step in the theory of social utility” that eventually culminated in the Bergson–Samuelson Social Welfare Function (SWF). The Bergson-Samuelson SWF presents collective social welfare as a function of the welfare of the individual members of the collective. The function converts each individual’s preference orderings over a given profile of alternative social states into a continuous number sequence that represents collective social welfare. Even though the function considers individuals’ preferences, it is classed as a ‘social’ welfare function because it accounts for each individual’s orderings over social states that impact on the welfare of all members of the collective.

Maximizing the value of the Bergson-Samuelson SWF, subject to the range of technically determined production possibilities, is intended to indicate policy arrangements which achieve Pareto optimality for the ethically preferred distribution of income, as well as for other factors assessed as socially significant. However, the validity of such outcomes was brought into question when the stability of individual’s preference orderings over finite social states was considered. Arrow (1951) investigated the social choice rules necessary to demonstrate that welfare is maximized when individual preferences are transformed into comprehensive and stable social welfare ordering. Arrow

introduced choice rules which, among other things, require preferences over different social states to be universal, so that orderings over any particular states are consistent with a complete set of profiles for ordering of preferences over all possible social states, and independent, so that social welfare orderings do not change in response to changes in social circumstances not included in the preference profile. In short, Arrow’s conditions require social welfare orderings to be universally path-independent.

Subject to these strong constraints, Arrow concluded that the only social welfare ordering that satisfies the Pareto principle is the ordering of a single dominant person. That is, when individuals’ ordinal preferences are the only data source from which judgments about social welfare are made, and judgments of this nature are made on the basis of transformed social welfare orderings, then it is impossible to make policy decisions on maximizing collective social welfare in terms of the Pareto principle for a democratic society. This result became known as the Arrow Impossibility Theorem. In a related manner, Sen (1970) wrote a seminal paper on the “Impossibility of the Paretian liberal”, which demonstrates that when a subset of society comprises non-liberal individuals, whose social welfare orderings suggest that their utility is increased by constraining others in at least one area of choice, then it is not always possible to introduce policy that is simultaneously liberal and Pareto optimal. To overcome these logical limitations of relying on preference orderings, there has been a partial return to utilitarian type analysis, with partial, and sometimes full, interpersonal comparability utility assumed

(Boadway and Bruce 1984) and reliance and on *quasi-Pareto* social improvements (Ng 1990).

Notwithstanding these limitations, social welfare theory remains relevant to public policy. It contributes to the logical development of policy by providing a basis for governments to commence the policy process with explicit consideration of the ethical precepts that it believe are necessary to maximize collective social welfare. Equally importantly, it assists critical evaluations of governments by enabling policy to be assessed in terms of the ethical precepts revealed by government conduct.

### **Selected References**

Albert, Michael and Hahnel, Robin. (1990) *A Quite Revolution in Welfare Economics*. Princeton: Princeton University Press.

Allais, Maurice. (1943) "Traité d'Economie Pure", in 1952 edition of *A la Recherche d'une Discipline Economique*. Paris: Imprimerie Internationale.

Arrow, Kenneth J. (1951) *Social Choice and Individual Values*. New York: Wiley.

Barone, Enrico. (1908) "The Ministry of Production in the Collectivist State", *Giornale degli Economisti*, as translated in F. Hayek (1935) (Editor), *Collectivist Economic Planning*. London: Routledge and Kegan Paul.

Bergson, Abram. (1938) "A Reformulation of Certain Aspects of Welfare Economics", *Quarterly Journal of Economics*, Volume 52, pp. 310-34.

Buchanan, James. (1962) "The Relevance of Pareto Optimality", *Journal of Conflict Resolution*, Volume 6, pp. 341-354.

Buchanan, James. (1987) "The Constitution of Economic Policy", *American Economic Review*, Volume 77, pp. 243-250.

Boadway, Robin and Neil Bruce. (1984) *Welfare Economics*. Oxford, Blackwell.

Chipman, John S. (1976) "The Paretian Heritage", *Revue Européenne des Sciences Sociales*, Volume 37, pp. 65-173.

Cullis, John and Jones, Philip. (1992) *Public Finance and Social Choice*. Maidenhead, McGraw-Hill.

Edgeworth, F.Y. (1881) *Mathematical Psychics*. London, Kegan Paul.

Groenewegen, Peter. (2003) "English Marginalism: Jevons, Marshal, and Pigou", in Warren J Samuels, Jeff E. Biddle and John B. Davis (Editors), *A Companion to the History of Economic Thought*. Oxford: Blackwell Publishing, pp. 246-261.

Hayek, F.A. (1935) *Collectivist Economic Planning*. London: Routledge and Kegan Paul.

Hicks, John. R. (1939) "The Foundations of Welfare Economics", *Economic Journal*, Volume 49, pp. 696-712.

King, J.E. (2003) "Non-Marxian Socialism" in *A Companion to the History of Economic Thought*, Warren J Samuels, Jeff E. Biddle and John B. Davis (Editors), Oxford: Blackwell Publishing, pp. 184-200.

Lange, Oskar. (1938) "The Economics of Socialism" in B.E Lippincott (Editor), *On the Economics of Socialism*. New York: McGraw-Hill, pp57-143.

Lange, Oskar (1942) "The Foundations of Welfare Economics", *Econometrica*, Volume 10, Number 3, pp. 215-228.

Lerner, Abba. P. (1932) "The Diagrammatical Representation of Cost

- Conditions in International Trade”, *Economica*, Volume 12, pp. 346-56.
- Lipsey, Richard G. and Kelvin Lancaster. (1956) “The General Theory of Second Best”, *Review of Economic Studies*, Volume 24, Number 1, pp.1-32.
- Ng, Yew-Kwang. (1990) “A Theory of Third Best”, in Yew-Kwang Ng (1990), *Social Welfare and Economic Policy*. New York: Harvester Wheatsheaf, pp.145-200.
- Pareto, Vilfredo. (1894) "Il Massimo di Utilità dato dalla Libera Concorrenza". in G. Busino (Editor), *Ouvres Complètes Tome XXVI, Écrites d'économie politique*. Genève: Librairie Droz, 1982, pp. 276-294.
- Pareto, Vilfredo. (1896-1897) *Corso di Economia Politica [Cours d'Économie Politique]*, Torino : Unione Tipografico-Editrice Torinese, 1971.
- Pareto, Vilfredo. (1909) *Manual of Political Economy [Manuel d'Économie Politique]* New York: Augustus M Kelly Publishers, 1971.
- Pareto, Vilfredo. (1913) “Il Massimo di Utilità per una Collettività in Sociologia”, in G. Busino (Editor), *Ouvres Complètes Tome XXII, Écrites sociologiques mineurs*. Genève : Librairie Droz, pp. 601-605, 1980.
- Pigou, Arthur C. (1920) *The Economics of Welfare*. London: McMillan.
- Samuelson, Paul A. (1954) “The Pure Theory of Public Expenditure” *Review of Economics and Statistics*, Volume 36, Number 4, pp. 387-389.
- Samuelson, Paul A. (1955) “Diagrammatic Representation of a Pure Theory of Public Expenditure”, *Review of Economics and Statistics*, Volume 37, Number 4, pp. 350-356.
- Scitovsky, Tibor. (1941) “A Note on Welfare Propositions in Economics”, *Review of Economic Studies*, Volume 9, pp. 77-88.
- Sen, Amartya. (1970) “The Impossibility of a Paretian Liberal”, *Journal of Politics and Economics*, Volume 78, Jan-Feb, pp. 152-7.
- Sidgewick, Henry. (1883) *Principles of Political Economy*. London: Macmillan.

Michael McLure  
Department of Economics  
University of Western Australia  
Perth, Australia  
michael.mclure@uwa.edu.au

## Policy Ineffectiveness Proposition

Jan Reijnders

### **Definition**

The policy ineffectiveness proposition posits that the government is incapable of managing output and employment in the economy. The proposition was introduced in the nineteen-seventies by Thomas J. Sargent and Neil Wallace (1975; 1976) as an extension of earlier work by Lucas (Lucas & Rapping 1969; Lucas 1972; 1973). It was the vehicle by which the concept of rational expectations was introduced into macroeconomic research and became one of the first exploits of the so-called New Classical counterrevolution in macroeconomic theory.

### **History**

With the policy ineffectiveness proposition, the New Classical School in fact returns to the principles of the Classical School which considered the economy as a self equilibrating system. The economy is assumed to automatically grind out optimal results as long as private initiative is left free to roam without being bothered by state intervention. This idea of '*laissez-faire*' as a guide to economic policy came under attack during the Great Depression of the 1930s. During this dramatic period, the free market economy proved to be incapable of maintaining full employment.

John Maynard Keynes ascertained that under conditions of *laissez-faire* wide fluctuations in employment can not be avoided and concluded that the duty of ordering the volume of investment that is necessary to produce full employment "cannot safely be left in private hands" (Keynes 1936:320). With his General

Theory Keynes lay the basis for the Keynesian Revolution that overturned the *laissez-faire* principle of economic policy. The latter was replaced by a view of a mixed economy where private initiative was still allowed to roam freely but where the stability of the system and the attainment of certain social goals such as full employment could only be guaranteed by government intervention. The Keynesian view that full employment could be maintained at all times by demand management became dominant after the Second World War.

A simplified version of the central idea is illustrated in Figure 1 which contains the confrontation of the aggregate demand (AD) and aggregate supply (AS) curve. The vertical axis represents the price level  $P$  and the horizontal axis represents output (which is positively related to employment). For a given level of autonomous (government) expenditure, the location of AD is determined by money supply  $M$ . It is downward sloping because with a given level of money supply (and given the level of autonomous expenditure), an increase in the price level implies a decrease in real money supply ( $M/P$ ). This pushes up the rate of interest. Because the rate of interest negatively affects aggregate demand, which in turn determines output, the level of output decreases when prices go up and increases when prices go down.

The Keynesian aggregate supply function (ASK) is upward sloping: An increase in output (and hence in employment) will put an upward pressure on nominal wages which in turn will push up prices. The ASK-curve only slopes gently because the increase in output is assumed to have only a limited effect on prices. The intersection between aggregate



demand and aggregate supply determines equilibrium point A. In the Keynesian view equilibrium output does not necessarily equal potential (that is full-employment) output. If A does not correspond to full

there is a permanent trade-off between inflation and unemployment, the Phillips-curve became one of the cornerstones of the Keynesian view on economic policy.

The Keynesian mainstream

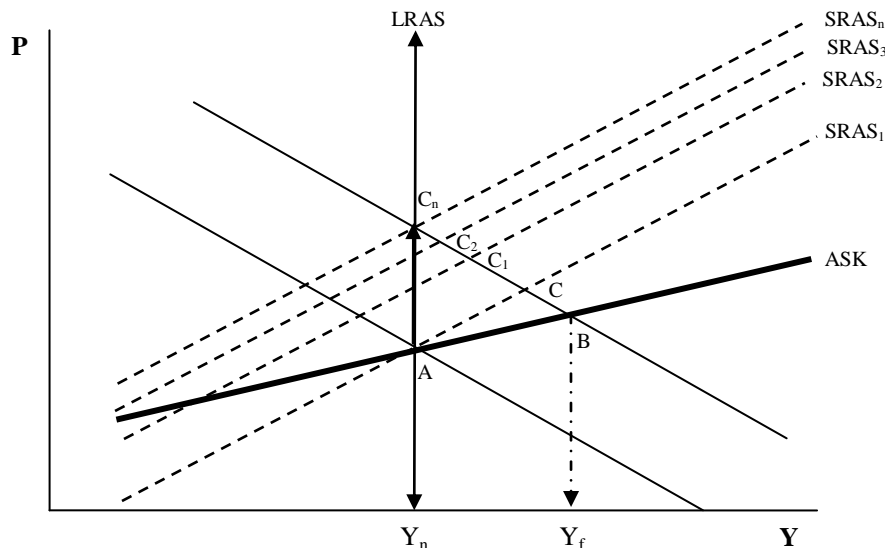


Figure 1: The Keynesian, Monetarist and New Classical view of the effect of a policy shock

employment there is scope for economic policy to increase output to its full employment level. This can be done by either monetary (increasing money supply) or fiscal policy (increasing government expenditure or reducing taxes). Suppose that monetary policy is used to increase money supply. Because the location of AD depends on total money supply, this increase shifts the AD-curve to the right. The result is a shift of the equilibrium position to point B that corresponds to a higher level of output and employment as well as prices.

On this view government can thus increase the level of employment and bring it closer to full employment if it is prepared to accept an increase in the price level. This is in line with the central idea of the Phillips curve, which relates to the negative link between the rate of change of the price level and the rate of unemployment (Phillips 1958). With its implication that

interpretation with the idea that an activist government could permanently influence the level of activity and was capable of pushing the unemployment rate towards zero, stayed in place until the second half of the 1960s. The acceleration of inflation that took place in this period sparked off what was dubbed the “Monetarist Counterrevolution”. Its main spokesman Milton Friedman challenged the central tenets of the Keynesian policy view. He argued that the unemployment rate could not be permanently pushed below a certain level which he designated the ‘natural rate of unemployment’ (Friedman 1968). In Friedman’s view the Keynesian proposition of a permanent inflation-unemployment trade-off is not satisfactory. It can only be upheld if one assumes that workers suffer from lasting money illusion. They are then assumed to base their labour supply decisions on nominal wages alone.

In actual fact, however, labour supply decisions are not based on nominal but rather on real wage considerations. In this case workers will seek to compensate any rise in the price level (which at a given nominal wage depresses their real wage) by a corresponding nominal wage increase. Accordingly expected price changes do play a role during wage negotiations. Friedman assumes that price expectations are formed by extrapolating recent price history (adaptive expectations). As a result tomorrow's nominal wage level comes to depend on today's (and/or yesterday's) price level.

In terms of the AS/AD diagram the following happens when the government increases money supply. Suppose that  $SRAS_1$  is the initial Monetarist (short run) aggregate supply curve (the curve is steeper than its Keynesian equivalent because in Friedman's world prices react stronger to changes in the level of economic activity). It is drawn on the assumption that the current price level corresponds to the expected price level ( $P_t = P_t^e$ ). Just like the Keynesian aggregate supply curve, the initial Monetarist aggregate supply curve intersects  $AD_0$  in equilibrium point A. The government's decision to increase money supply leads to an upward shift of the AD-curve (to  $AD_1$ ). Now  $C_1$  becomes the new equilibrium point. Unlike the Keynesian equilibrium at B, point  $C_1$  is only transitory.

The short run aggregate supply function is built on the assumption that equilibrium is defined by the equality of actual and expected price levels. Therefore the change in the actual price level will increase the expected price level which implies that the next short run aggregate supply curve will incorporate this change. Accordingly the new short run aggregate supply curve will

be situated to the northwest of the previous one ( $SRAS_2$ ). Since the new short run equilibrium again implies a still higher price level, the expected price level will be adjusted in an upward direction. As a consequence of this the aforementioned process repeats itself with the result that  $SRAS_3$  is once again situated at a higher level, and so on and so forth. The short run equilibria move up along the  $AD_1$ -curve. In the process the gap between actual and expected price level becomes smaller and smaller until expectations finally catch up with the actual price level at point  $C_n$ . It is here where process stops and the economy reaches its long run equilibrium point.

The upshot of all this is that the original monetary impulse—which initially led to an increase in the real level of economic activity—is gradually eaten away by the price rise because the latter reduces real money supply, pushes up the interest rate, reduces aggregate demand and hence curbs output and employment. Accordingly the monetary impulse initially has a real effect, a move from output  $Y_0$  to output  $Y_1$ . But this is only temporary because economic activity in the end always returns to its initial level at  $Y_0$ , that is what Friedman called its “natural” level. Therefore Friedman's long-run aggregate supply curve will be a vertical line at point  $Y_0$  (LRAS in Figure 1).

Economic policy is only capable of pushing activity beyond this line for a short period of time. Any effort to permanently surpass the natural rate will necessarily involve a permanent increase in nominal money supply which results in a ever accelerating inflation. Accordingly there can be no such thing as a permanent trade-off between unemployment and inflation as the Keynesians thought. There is only a short run trade-off as implied by what

Friedman calls the “expectations augmented” Phillips curve.

In doing so Friedman made a distinction between the long run and the short run. In the long run equilibrium is completely supply-determined. This makes it resistant to policy. In the short run, however, demand shocks can cause deviations from long run equilibrium. This gives an edge to demand management and creates a time frame wherein economic policy can be effective. The window of opportunity for policy is determined by the stretch of time that expectations need to adjust to the actual price level. By his long run / short run dichotomy Friedman appeared to have reached a compromise wherein the classical idea of a supply determined equilibrium is squared with part of the Keynesian notions of disequilibrium. In the short run the economy is allowed to exhibit demand driven, short run fluctuations around a long run supply determined equilibrium trend.

Before long the latitude for fluctuations and the concomitant elbow room for economic policy was further reduced by Lucas’s criticism of Friedman’s treatment of the process of expectation formation. In Lucas’s view adaptive expectations are fundamentally flawed for two reasons. The first is that adaptive expectations are backward looking. This does not appear to be adequate in the context of expectations, which - by their very nature - are supposed to be forward looking. The second is that adaptive expectations imply that economic agents only gradually adjust to predictive errors from the past (in terms of Figure 1, they adjust their expectations step by step in a number of rounds before reaching the final equilibrium stage).

This implies that agents err systematically and don’t learn from their

own mistakes, a phenomenon that can hardly be squared with the common assumption of a rational *homo economicus*. Instead Lucas proposes to model expectations in the manner suggested by John Muth dubbed “rational expectations”(Muth 1961). Rational expectations are explicitly forward looking. They are supposed to coincide with the optimal economic forecasts which use all available information and which are essentially the same as the predictions made by the relevant economic theory given the stochastic properties of the available data. With rational expectations, economic agents learn from their own mistakes and consequently don’t err systematically.

In the case of Figure 1, the rationally expecting economic agent would know the initial position. Provided that the monetary impulse could not be foreseen, the short run aggregate supply curve would shift to the right to constitute a new short run equilibrium point at  $C_1$ . After the fact the agent would, however, immediately recognize the rightward shift to  $SRAS_1$  and the resulting short run equilibrium  $C_1$ . From economic theory she would then immediately predict that the new long run equilibrium would be at  $C_n$ . So in stead of Friedman’s cumbersome walk across the shifting  $SRAS_2 \dots SRAS_3$  sequence, she would immediately jump from  $C_1$  to  $C_n$ . In this way Lucas maintains Friedman’s long run short run dichotomy.

There is still scope for a real influence of monetary policy, be it that the window of opportunity has been narrowed considerably. Firstly because the time span within which policy is supposed to exert its influence is reduced significantly (the time span of adjustment is in fact abridged to almost zero). Secondly because the policy

action is only effective if it is not anticipated by private agents. From this perspective monetary policy becomes a kind of strategic game wherein government measures are only effective if they take private agents by surprise (for this reason Lucas's model revolves around the so-called "surprise supply function"). Foreseeable policy measures lead to an immediate adjustment of price expectations (i.e. a jump from point A to  $C_n$ , as indicated by the arrow in Figure 1) which leaves the real activity level as it is.

### **Theoretical Debate**

With respect to the introduction of rational expectations in the macroeconomic model, the sting is in the tail: it appears that policy is only effective if it is not foreseen. This issue is taken up by Sargent and Wallace who make an effort to Figure out how rational expectations in the context of the natural rate hypothesis work out on economic policy.

In the context of an ad hoc macroeconomic model they compare two alternative strategies for the monetary authorities. One is to peg the interest rate period by period and adjust money supply accordingly. The other is to set money supply period by period while letting the interest rate be whatever the market brings about. They then study the effects of these alternative strategies first on the basis of an autoregressive model in which public expectations are formed adaptively, and second on the basis of a model based on rational expectations. The reason for comparing the two alternative strategies is to do justice to the differences between the Monetarists and the Keynesians. Monetarists are in favour of a simple rule without feedback (i.e. growth of the money supply with a fixed percentage per period)

whereas the Keynesians adhere to the view that monetary policy must have a feedback rule which creates the possibility to counter the effects of external disturbances in order to attenuate the business cycle.

The authors first demonstrate that the optimal monetary rule must have a feedback rule provided that the response of the system to shocks leads to persistence (that is serially correlation of successive values of the goal variable). This seems to corroborate the Keynesian position. It is observed, however, that the Keynesian view is implicitly based on the following basic propositions: First that the economic structure is characterised by extensive simultaneity, second that, due to time-lags, the effect of shocks is distributed over time, and third that the lag structure is constant over time. Particularly the third proposition is crucial in this context. Lucas argues that the underlying assumption is inappropriate (Lucas 1976). It pretends that the economy and accordingly the parameters of the model remain constant over time.

However, if rational agents are aware of the policy rule (either because monetary authorities have announced its structure or because the agents are able to infer it from recent experience) they will incorporate this rule when forming expectations (Muth rationality). In this case policy changes behaviour and accordingly the parameters of the model change as well. To do justice to this so-called 'Lucas critique', Sargent and Wallace incorporate the policy rule in the model by including it in the formula that expresses the formation of expectations. They then demonstrate that the reduced form of this model no longer contains information on the parameters of the policy rule. Accordingly the outcomes of the model become completely

independent of economic policy. Economic policy can only have a real effect if the actual setting of the policy instrument differs from the setting expected by the public. This is only during the period when the public is taken by surprise and has not yet been able to adjust to the different settings of the policy instrument. In the context of a Philips-curve model Sargent and Wallace then compare the results of including different specifications for the process of expectation formation. Again it turns out that economic policy is only effective under conditions of adaptive expectations. Only in this case are the authorities capable of influencing the rate of unemployment by 'fooling' the people (Hall 1976± 141). In practice, however, it is not possible to trick the public all the time.

As soon as the assumption is introduced that people act 'rationally' in the sense that their expectations are equal to the objective mathematical expectations conditional on the information known at the time that the expectation is formed, the effect of monetary policy vanishes. Under these circumstances the monetary authorities have no option at all to conduct countercyclical policy: "Thus combining the natural rate hypothesis with the assumption that expectations are rational transforms the former from a curiosity with perhaps remote policy implications into a hypothesis with immediate and drastic implications about the feasibility of pursuing countercyclical policy" (Sargent and Wallace 1976 :178).

The introduction of this phenomenon, dubbed the 'policy ineffectiveness proposition' gave rise to an extensive debate. The critics initially focused on the theoretical foundation of the proposition

and in a later stage on the robustness of the results of the empirical tests.

The theoretical criticism derived from the observation that the basic theorem critically depends on the combination of two key elements: Rational-expectations (in the Muthian sense) and perfect wage and price flexibility with instantaneous market clearing. Rational expectations in the Muthian sense presuppose that all economic agents command an enormous amount of information and use it effectively, that they know the 'true' model of the economy, that they have accurate estimates of all parameters of this model, that they have sufficient information to identify the systematic components of economic policy and are capable of adjusting immediately to changes thereof. It is argued that such enormous amounts of information cannot possibly be processed by human beings (not even by the representative agents that man the world of the New Classics).

But even accepting this, there is a problem when account is taken of the assumptions made when configuring the basic model of the new classical approach. It is argued that policy ineffectiveness is actually introduced by definition. The combination of the assumption that only price or wage surprises cause deviations from the natural rate with the assumption of rational expectations in the Muthian sense, that is that the deterministic part of government behaviour is included in the information sets that determine private forecasts, inevitably leads to policy ineffectiveness (Shiller 1978: 13; Buiter 1980: 39; also see: Farmer and Geanakoplos 2008). Contrary to this there are ample theoretical as well as practical reasons for assuming that certainly fiscal policy but also monetary policy will have

real effects even if they are fully anticipated (Shiller 1978:12; Buiter 1980).

Another set of reasons why policy ineffectiveness can hardly be squared with reality has to do with the assumptions regarding perfect price and wage flexibility with instantaneous market clearing. In practice the existence of long term wage contracts and the fact that they expire at different points in time (staggering of wage decisions) introduce nominal rigidities that result in wage and price stickiness. If it takes time for wages and prices to adjust, markets don't clear instantaneously and there is scope for economic policy. Fisher, Phelps and Taylor convincingly demonstrated that in economic models combining rational expectations with sticky wages and prices government policy renders real effects even if it is fully anticipated (Fischer 1977; Phelps & Taylor 1977; Fischer 1979; Taylor 1979; Taylor 1979; Taylor 1984; Taylor 1998).

### **Empirical Debate**

To a certain extent, the theoretical discussion was a dialogue of the deaf since the proponents of the policy ineffectiveness proposition maintained that it is not admissible to test a theory by the realism of its assumptions. It is only the accuracy of its predictions that counts (Friedman 1964). Accordingly the decisive battle over the policy ineffectiveness proposition was fought over the accuracy and robustness of its empirical tests.

Contrary to the expectations of some of the critics (Taylor 1989: 186) the policy ineffectiveness proposition was more than a flash in the pan. Following earlier efforts of Lucas and Sargent to provide adequate means for testing its central tenets (Lucas 1972; Sargent 1973), the policy ineffectiveness proposition got special

impetus after the publication of Robert Barro's empirical tests which appeared to corroborate Sargent and Wallace's findings (Barro 1977; Barro 1978; Barro 1979; Barro & Rush 1979; Leiderman 1980). The author set out to present a rigorous test of the hypothesis that "only unanticipated movements in money affect real economic variables like the unemployment rate or the level of output" (Barro 1977: 101).

The results are obtained in a two-stage procedure. In the first stage Barro uses regression analysis to fit a simple model of money growth in order to determine the anticipated part of actual money growth in the USA between 1941 and 1973. Since the fitted model is supposed to represent the anticipated part of money growth, the residuals from this regression are consequently considered to represent the unanticipated part. This unanticipated part is subsequently inserted in the set of independent variables for the second stage regression that is used to explain the movements of the unemployment rate over the period 1946-1973. It turns out that unanticipated money has a significantly negative impact on the unemployment rate in the mentioned period.

In a sequel to his 1977 paper, Barro extends this analysis to the domain of output and prices. In his view the results corroborate his earlier findings in the sense that the analysis between output and unanticipated money growth conform to his earlier results on unemployment. Likewise, the results for price level movements are consistent with the central hypothesis of a contemporaneous one-to-one correspondence between anticipated money and the price level (Barro 1978: 550).

The empirical tests met with several types of criticism. The first strand of these

relates to methodology in general. It is pointed out that the tests as proposed by Lucas (1972), Sargent (1973) as well as Barro (1977) and their direct descendants have in common that they test the conformity of the implications of their hypothesis with the available evidence without considering a possible competing hypothesis. As argued by Pessaran and Gordon, for instance, sound methodology requires that at least one genuine alternative hypothesis is considered. The latter fill in this lacunae and subsequently demonstrate that the policy ineffectiveness proposition does not survive the confrontation with viable alternatives, that is hypotheses that have the characteristic traits of the (Neo-)Keynesian model (Gordon 1982; Pesaran 1982).

A second strand of criticism adduces that Barro's results directly depend upon the way in which his tests are specified. Accordingly it is argued that his results are direct descendents of the assumption that the supply function is linear (Shiller 1978: 10) or the assumption that only short lags are involved in the output and unemployment assumptions (Mishkin 1983:127). A third strand of criticism implies that Barro's decomposition of money growth in its anticipated and unanticipated components is questionable. In Barro's two-stage procedure, the findings of the first stage (the test of the rational expectations hypothesis) completely determine what is left to be found in the second stage (the test of the unanticipated money hypothesis). But because the nature of the underlying theorem, the two hypotheses should be tested jointly. Application of the appropriate methodology, however, inevitably leads to results that are rather damaging to the view that the policy

ineffectiveness hypothesis is supported by the empirical evidence (Mishkin 1982; Mishkin 1982; Pesaran 1982; Mishkin 1983; Pesaran 1988).

In view of these destructive results it must be concluded that the proponents of PIP ultimately lost the battle over the empirical testing, which is crucial in view of the Friedman/Lucas type positivist methodology they adhere to.

### **Demise of the Proposition**

It is obvious that the policy ineffectiveness proposition can no longer be upheld when it turns out that the addition of a touch of realism to the model in the shape of wage or price stickiness (as demonstrated by Phelps, Fischer and Taylor), a small extension and working up of its specification (as demonstrated by Pesaran and Mishkin), as well as the confrontation with a viable alternative (as demonstrated by Gordon and Pesaran) each turn out to have destructive effects on the proposition. Barro's influential paper of 1977 and its direct descendants (Barro 1978; Barro 1979; Barro & Rush 1979; Leiderman 1980) kept the debate going for a considerable amount of time but ultimately the evidence in favour of the policy effectiveness proposition was completely overwhelmed by the empirical evidence against this view. In the end even Barro himself had to give up on the proposition when he found strong evidence against it (Barro & Hercowitz 1980).

### **Selected References**

Barro, R.J. (1977, March) "Unanticipated Money Growth and Unemployment in the United States", *American Economic Review*, Volume 67, Number 1, pp. 101-115.

- Barro, R.J. (1978) "Unanticipated Money, Output, and the Price Level in the United States", *Journal of Political Economy*, Volume 86, Number 4, pp. 549-80.
- Barro, R.J. (1979) "Unanticipated Money Growth and Unemployment in the United States: Reply", *American Economic Review*, Volume 69, Number 5, pp. 1004-09.
- Barro, R.J. and Z. Hercowitz. (1980) "Money Stock Revisions and Unanticipated Money Growth", *Journal of Monetary Economics*, Volume 6, Number 2, pp. 257-67.
- Barro, R.J. and M. Rush. (1979) *Unanticipated Money and Economic Activity*. NBER Working Paper No W0339. Cambridge, Mass: NBER.
- Buiter, W.H. (1980) "The Macroeconomics of Dr. Pangloss: A Critical Survey of the New Classical Macroeconomics", *Economic Journal*, Volume 90, Number 357, pp. 34-50.
- Farmer, J.D. and J. Geanakoplos. (2008) *The Virtues and Vices of Equilibrium and the Future of Financial Economics*. Cowles Foundation Discussion Paper 1647. New Haven: Cowles Foundation, Yale University.
- Fischer, S. (1977) "Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule", *Journal of Political Economy*, Volume, Number 1, pp. 191-205.
- Fischer, S. (1979) "Anticipations and the Nonneutrality of Money", *Journal of Political Economy*, Volume 87, Number 2, pp. 225-52.
- Friedman, M. (1964) "The Methodology of Positive Economics", in Milton Friedman, *Essays in Positive Economics*. Chicago & London: University of Chicago Press: 3-43.
- Friedman, M. (1968) "The Role of Monetary Policy", *American Economic Review*, Volume 58, Number 1, pp. 1-17.
- Gordon, R.J. (1982) "Price Inertia and Policy Ineffectiveness in the United States, 1890-1980", *Journal of Political Economy*, Volume 90, Number 6, pp. 1087-1117.
- Hall, R.E. (1976) "The Phillips Curve and Macroeconomic Policy", *Carnegie Rochester Conference Series on Public Policy*, 1976, pp. 127-48.
- Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*. London & Basingstoke, MacMillan.
- Leiderman, L. (1980) "Macroeconometric Testing of the Rational Expectations and Structural Neutrality Hypotheses for the United States", *Journal of Monetary Economics*, Volume 6, Number 1, pp. 69-82.
- Lucas, R.E., Jr. (1972) "Econometric Testing of the Natural Rate Hypothesis", in O. Eckstein (Editor), *The Econometrics of Price Determination*. Washington DC: Publications Services, Division of Administrative Services, Board of Governors, Federal Reserve System: pp. 50-59.
- Lucas, R.E., Jr. (1972) "Expectations and the Neutrality of Money", *Journal of Economic Theory*, Volume 4, Number 2. pp. 103-24.
- Lucas, R.E., Jr. (1973) "Some International Evidence on Output-Inflation Tradeoffs", *American Economic Review*, Volume 63, Number 3, pp. 326-34.
- Lucas, R.E., Jr. (1976) "Econometric Policy Evaluation: A Critique", *Carnegie-Rochester Conference Series on Public Policy* 1976, pp. 19-46.



- Lucas, R.E., Jr and L.A. Rapping (1969) "Real Wages, Employment, and Inflation", *Journal of Political Economy*, Volume 77, Number 5, pp. 721-54.
- Mishkin, F.S. (1982) "Does Anticipated Aggregate Demand Policy Matter? Further Econometric Results", *American Economic Review*, Volume 72, Number 4, pp. 788-802.
- Mishkin, F.S. (1982) "Does Anticipated Monetary Policy Matter? An Econometric Investigation", *Journal of Political Economy*, Volume 90, Number 1, pp. 22-51.
- Mishkin, F.S. (1983) *A Rational Expectations Approach to Macroeconometrics. Testing Policy Ineffectiveness and Efficient-Markets Models*. Chicago & London: University of Chicago Press.
- Muth, J.F. (1961) "Rational Expectations and the Theory of Price Movements", *Econometrica*, Volume 29, Number 3, pp. 315-335.
- Pesaran, M.H. (1982) "A Critique of the Proposed Tests of the Natural Rate-Rational Expectations Hypothesis", *Economic Journal*, Volume 92, Number 367, pp. 529-54.
- Pesaran, M.H. (1988) "On the Policy Ineffectiveness Proposition and a Keynesian Alternative: A Rejoinder", *Economic Journal*, Volume 98, Number 391, pp. 504-08.
- Phelps, E.S. and J.B. Taylor. (1977) "Stabilizing Powers of Monetary Policy under Rational Expectations", *Journal of Political Economy*, Volume 85, Number 1, pp. 163-90.
- Phillips, A.W. (1958) "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957", *Economica (New Series)*, Volume 25, Number 100, pp. 283-299.
- Sargent, T.J. (1973) "Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment", *Brookings Papers on Economic Activity*, Number 2, pp 429-472.
- Sargent, T.J. and N. Wallace. (1975) "'Rational' Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule", *Journal of Political Economy*, Volume 83, Number 2, pp. 241-54.
- Sargent, T.J. and N. Wallace. (1976) "Rational Expectations and the Theory of Economic Policy", *Journal of Monetary Economics*, Volume 2, Number 2, pp. 169-83.
- Shiller, R.J. (1978) "Rational Expectations and the Dynamic Structure of Macroeconomic Models: A Critical Review", *Journal of Monetary Economics*, Volume 4, Number 1, pp. 1-44.
- Taylor, J.B. (1979) "Estimation and Control of a Macroeconomic Model with Rational Expectations", *Econometrica*, Volume 47, Number 5, pp. 1267-86.
- Taylor, J.B. (1979) "Staggered Wage Setting in a Macro Model", *The American Economic Review. Papers and Proceedings*, Volume 69, Number 2, pp. 108-113.
- Taylor, J.B. (1984) *The Role of Expectations in the Choice of Monetary Policy*. NBER Working Paper 1040. Cambridge, Mass: NBER.
- Taylor, J. B. (1989) "The Evolution of Ideas in Macroeconomics", *Economic Record*, Volume 65, Number 189, pp. 185-89.

- Taylor, J.B. (1998) *Staggered Price and Wage Setting in Macroeconomics*. NBER Working Paper 6754. Cambridge, Mass: NBER.
- Wade, J. (1833) *History of the Middle and Working Classes*. London: E. Wilson.
- Yule, G. U. (1927) On a Method of Investigating Periodicity in Disturbed Series with Special Reference to Wolfer's Sunspot Numbers. *Philosophical Transactions of the Royal Society of London*, (A 226), pp. 267-298.
- Zarembka, P. (2003) (Editor) *Value, Capitalist Dynamics, and Money*. Amsterdam: Reed Elsevier.

*Jan Reijnders*  
*School of Economics, Utrecht University*  
*Janskerkhof 12, Utrecht, Netherlands*  
*j.reijnders@uu.nl*

## Political Business Cycles

*Jerry Courvisanos*

### **Definitions and Context**

The term ‘business cycles’ refers to the recurring expansion and contraction of economic activity. Employment of labour is the most direct and identifiable outcome of the cyclical activity in an economy. Employment rises due to expansion through improved economic performance as a result of higher aggregate demand for products. With a short time lag employment then falls due to contraction as lower aggregate demand reduces the need for workers. Business cycles in capitalist economies have a periodicity (peak-to-peak) of around eight to twelve years, as a broad generalisation.

The causes of business cycles has been an extensive research project for many economists ever since the rise of the industrial revolution, with much effort being placed on mechanisms that generate such recurring (if not regular) cycles of activity (Zarnowitz 1985). Such analyses identify the endogenous and exogenous variables that influence this cycle pattern in models of pure capitalist economies without any government intervention. From a Keynesian (endogenous) perspective, these cycles are self-sustaining and driven by the investment spending decisions on capital goods by firms. From the neo-classical (exogenous) perspective, these same cycles are responses to monetary and real disturbances.

Political business cycles (PBCs) refer to the additional layer of business cycle analysis over the pure business cycle that comes into operation with government intervention. Essentially, these are the political causes of any business cycle

patterns that occur over and above the strictly economic causes. With government intervention of any kind the PBC comes into play. In the early part of the twentieth-century, prior to Keynesian economic analysis, a balanced budget with no monetary intervention was considered to have no political implications for the business cycle. British ‘mandarin’ economist, John Maynard Keynes, introduced the idea of effective demand as the major factor on the business (or ‘trade’) cycle, which implies that any balanced budget has a stronger expansionary multiplier effect through direct government spending, compared to the contractionary multiplier effect of tax collections. This is due to the Keynesian ‘fact’ that taxation reduces personal consumption spending (excluding taxation out of savings) by less than what government spending increases effective demand. Thus, by governments making a decision to intervene in the economy by spending and taxing, this sets up a PBC path.

Governance of the business cycle path becomes a crucial issue for any government in a capitalist economy, both from the perspective of winning (and retaining) electoral power and in terms of responsible economic managers. The central question is how does public policy affect the path of economic activity and what are the implications for unemployment, inflation, and innovation of the economy? Answers to this question depend on the type of PBC that is specified. In the literature there exists the class-based PBC and the electoral-based PBC, with two versions of each type-base.

### **Class-Based Political Business Cycles**

The first rigorous theory on the PBC is credited to Michał Kalecki in his

extensively quoted article, *Political Aspects of Full Employment* (1943). PAFE has a Marxian class analysis as the foundation, where the capitalist class prevails over the political institutions of the society. Kalecki (1971) is a reprint of selected essays, including PAFE to which Kalecki notes in the Preface as having correct predictions “*grosso modo*”. This statement leaves class-based PBC open to two interpretations: stop-go ‘fine tuning’ and longer-term ‘political trend’. The former is applied to the variations in cycle amplitude, especially during the relatively less unstable ‘Golden Period’ of post-war reconstruction and mass consumption that ended with the first oil price shock in 1973. The latter is applied to the change in political priority from full employment (anti-unemployment) in the 1945-73 (Golden) period to price stability (anti-inflation) since 1973.

Boddy and Crotty (1975) and Robinson (1976) develop the former version in terms of the stop-go Keynesian policies of the ‘Golden Period’ that are aligned with the cyclical profit squeeze of capitalist firms. Profits begin to be squeezed at the top of the boom as consumption spending slows down. Capitalist entrepreneurs have the exclusive control of capital investment, and under these profit squeeze conditions they make the decision to reduce their investment spending which turns the pure capitalist business cycle into a contraction phase. Keynesian macroeconomic policies provide the method by which this contraction can be prevented and full employment be maintained. This is through stimulation of effective demand *via* government spending, reduction of taxes and lowering interest rates (with increased money supply). The three policy

instruments can be employed to affect an expansionary impact on the economy.

Keynes (1936) expressed the faith that the power of this effective demand idea will be used to stabilise the business cycle at full employment. Kalecki (1943) too recognised the powerful tool of effective demand management by government (since he developed this same analysis in 1933), but he also recognised the stronger power of business interests to prevent this from occurring. Kalecki specified three fears that capitalists have with Keynesian full employment: (i) Loss of *economic* control with businesses’ state of confidence reflected in investment decisions and the business cycle that they generate; public policy demand management would effectively deprive capitalists of their power to influence economic conditions and also governments themselves; (ii) Loss of *policy* control as governments extend their impact through their own investment spending into the areas regarded as business’s legitimate sphere of influence (e.g. transport, public utilities); (iii) Loss of *industrial* control of the workforce if governments are able to maintain full employment over the long-run, so that ‘the sack’ ceases to play its disciplinary role for businesses.

Towards the top of the expansion phase of the cycle, the combination of profit squeeze and inflationary pressures manifests itself in a significant negative shift to the state of business confidence. This is reflected in profit rates falling, financial gearing rising and capacity utilisation falling as large capital investment projects come on stream at the time when consumption rates are slowing down (Courvisanos 1996). Business interests enlist rentiers interests to support them in having mainstream economists

identifying the economy as “unsound” (Kalecki 1971:144). Pressure is placed on governments to renege on full employment commitments and introduce the ‘stop’ elements of fine tuning by dampening effective demand policy instruments. This ensures the demise of old capital stock and the reduction in real wages, essential in the renewal of capitalism.

The ‘go’ policy elements of government stimulation in effective demand are then used when business interests enlist workers to support them in having mainstream economists declare a slump as detrimental to the economy. There would, however, be strong debate between all these supporters of stimulation as to the precise instruments and extent of their use. A stimulation package allows the cycle to move into a new expansionary phase, with new capital stock coming forward on the basis of innovation in newer technological developments. Old capital stock can then be decommissioned so that utilisation rates are manageable in relation to new investment spending (Galbraith & Darity, 1994:459-68). Kennedy (1973) provides empirical support in the U.K (1953 to 1971) for a “predominantly” planned ‘stop-go’ policy approach, with technical errors due to poor forecasting playing only a relatively minor role.

Steindl (1979), Bhaduri and Steindl (1983) and Catley and McFarlane (1981) use PAFE to explain the long-term implications of the PBC in terms of a ‘political trend’. These studies draw on the historical development of advanced capitalist economies like USA, U.K. and Australia to show that the shift in economic policies in the early 1970s from Keynesian ‘stop-go’ policies to Friedman monetarism and neoliberalism is due to the same three fears Kalecki identified in PAFE. The

difference is that in this version of the class-based PBC a longer timeframe allows for what Mair and Laramie (2002) refer to as “feedbacks between capitalists and workers over the political and social tensions of full employment to work themselves through.” These feedback effects generate rent seeking behaviour by powerful monopoly control interests who form ‘distributional coalitions’ to shift profit shares upwards by establishing obstacles in the road to full employment. These coalitions reduce efficiency and depress the adoption of new technologies in an effort to skew the income shares. Mair and Laramie (2002) provide empirical evidence to reveal the end of the post-war ‘full employment’ stop-go strategy in the early 1970s coincided with the only significant period of income share turbulence. Aschauer (2000) sets out empirical evidence for the USA that supports this contractionary political trend with the decline of public investment since the early 1970s. Catley and McFarlane (1981) provide similar historical evidence for Australia.

The contractionary political trend enabled the capitalist class to assert its economic and social dominance over labour and to cleanse capital of inefficient and oversupplied old stock. This process varies in time over different economies, with the USA leading the way after the 1990-92 recession into a new age of active innovation, stimulating large private investment spending and generating a new expansionary political trend. This expansionary trend consists of a new dynamic in public policy governance that has a PBC perspective as governments support re-armament (and war), tax cuts for the rich, innovation-supported subsidies and allowances, and ‘cheap money’ in a

fragile financial system. Kalecki (1945) identified these stimulatory policies, and they have now been reactivated by the USA at the start of the 21<sup>st</sup> Century.

The class-based PBC provides a crucial appreciation of the way governance of the business cycle operates and also the challenges for public policy implementation in the face of capitalist interests. In this respect some aspects of the PBC approach are inadequate and need to be developed further. The areas of most concern are greater financial instability, market-based globalization and mobility of capital funds, shifting of macroeconomic policy to central bank control with balanced (or surplus) budgets, and shift in politics from 'right to work' to 'right to manage' and the corporate governance problems that this has engendered.

### **Electoral-Based Political Business Cycles**

PAFE has an element of electoral concern that faces all capitalist democracies, when it notes the necessity that "something must be done in the slump" to stimulate the economy. Nordhaus (1975) acknowledges PAFE as the "only serious theory" on PBCs, then creates an electoral-based version of the PBC that is driven by politicians who manipulate macroeconomic instruments in concert with the electoral policy cycle. This type of PBC removes the class-base of PAFE and significantly shifts the initiating force from Kalecki's "business interests" to politicians with an eye to getting elected in the upcoming election and the aftermath of the election.

The electoral-based PBC developed a separate and dominating position in the economics literature as mainstream economists now had a PBC analysis that could be accommodated inside the standard neo-classical framework where the

exogenous political influence on the business cycle could be divorced from the role of market-based business. Political scientists have proceeded to distinguish between two versions of this electoral-based PBC: (i) electoral vote-maximising model with office-seeking policy makers and (ii) partisan vested interests model with policy-seeking policy makers. A vast political science literature has grown around these two versions and been recently reviewed in great detail by Franzese (2002).

The electoral (vote-maximising) business cycle is the version in which the election cycle is the basis of politicians garnering popular support on the assumption that the general populace takes macroeconomic conditions of the business cycle into account when voting. Wright (1974) and Nordhaus (1975) are the initial authors of this version. As an election approaches, the incumbent government tries to ensure its re-election by expansionary policies that serve two purposes. First, the popularity of the government is increased as employment and real incomes rise. Second, the financial and business indicators will be interpreted in the media as forecasting continued boom conditions as the voters approach the ballot box. After the election, whichever party wins power, the government needs to react to the concerns of the financial press and business that the boom conditions will create unsustainable inflationary pressures. The populace begins to be concerned that any gains in the boom will be erased as price rises reduce real incomes. Newly elected governments respond through contractionary policies that dampen the economy and set the business cycle into contraction phase. The 'toughness' of governments in 'making the hard

decisions' is applauded in the media, while the government has time until the next election in three to four years time creates the circumstances for expansionary policies that repeats the politico-economic cycle (Frey 1978). Public policy theory lends support to this PBC version.

Empirical evidence is hard to generate that can fit into this electoral business cycle thesis. Many attempts have been made with "negative and inconclusive" results (see, for details, Osiatyński 1990:575). Theoretical limitations also exist to this thesis in terms of Kalecki's framework. Intervals of about four to five years between elections does not leave enough time for the disciplinary role of unemployment to act on unions and workers in bargaining for wages, nor enough time to effectively reduce the role of the State. There is an electoral connection to any business cycle, and this model provides a basis for understanding the processes operating (Tufte 1978). However, governments do act in ways that often seem to be in contradiction to this electoral process (see Galbraith and Darity 1994:463-4) and this often has to do with responding to pressures from particular vested interests. This concern has led to the development of the second version of the electoral-based PBC.

The partisan business cycle thesis places vested interests within the context of the electoral cycle. Hibbs (1977) first developed this version, and it was refined by Alesina (1989) and by other Alesina articles with different collaborators. In this version, politicians are partisan in the way they deliver benefits to their constituent vested interests. The timing of economic policy activity is based on expansionary-based selected incentives to particular constituencies coming up to an election

and then shifting to contractionary-based incentives to satisfy different constituencies after the election. The assumption is that voters react in a partisan way to politicians' incentives that creates business cycle patterns.

Empirical studies on evidence of the partisan cycle thesis are related to tracking the timing of the electoral calendar with correlation to the incumbent partisan constituencies that are being influenced. Empirics reviewed by Franzese (2002) create a dilemma for this version. On the one hand, the evidence reveals strong partisan shifts that effect real economic performance, but not necessarily in tune with the electoral cycle. The electoral policy cycle does however exhibit partisan shifts within specific economic policies (monetary, fiscal, innovation, competition etc.), but this does not have the same significant effect on economic activity as altering the aggregate effective demand in the class-based PBC. General equilibrium neo-classical approaches incorporating rational expectations into the political behaviour of governments and interests groups, called rational partisan cycles, have been inadequate in explaining the empirical patterns observed. Yet, Franzese (2002:369) concludes that although relatively new in approach, "research into such context-conditional electoral and partisan cycles seems to offer much promise for resolving anomalies and an ideal substantive venue for theoretical and empirical advancement..." in the study of PBCs.

### **Governance Issues**

The PBC as a concept can be useful if it provides understanding of past governance practices and assists in managing future governance issues in the face of inherent

economic fluctuations. Irrefutable statistical proof of any particular version of the PBC does not exist in the literature (see Frey 1997). At the theoretical level, increasingly specialised research has diluted the original notion of the PBC. Specific mechanistic models developed for statistical validation are inappropriate in the fluid world of politics. What is left is a set of well-formulated, but divergent, political analyses that explain in various ways how pure capitalist-based economic fluctuations are exacerbated through political machinations.

All PBC versions can contribute valid elements to an overall PBC framework for governance issues. The PBC framework requires a class-based foundation that is modified by the electoral needs of politicians and partisan needs of vested interests. Such a framework would have to be underpinned by three crucial politico-economic features that motivate the PBC concept. These three features are:

‘Stop-go’ economic policies have largely given way to ‘political trend’ policies relating to market-oriented economic commitments; together with discretionary political power in foreign wars, financial regulation and international relations.

Macroeconomic policy variables (i.e. employment, price stability) that affect all voters as against other policy variables that have relatively greater variability of affect between different groups of voters.

Limits to achieving full employment through economic stabilisation policies as capitalism demands renewal of capital stock through technological innovation via boom-bust cycles of economic activity.

These features are common to all advanced capitalist economies to which PBCs have been investigated; only the

extent that these features affect various nations differs. The English-speaking nations are most influenced by the market-orientation political trend, while Scandinavian nations and Japan have a stronger employment commitment. Mid-European nations like Germany, The Netherlands and France place more emphasis on economic stabilising elements. Southern European nations are still significantly subject to ‘stop-go’ policies due to less sophisticated economic and political institutions. Even nations without representative western democracies have shown PBC patterns subject to partisan and class-based interests. See Frey (1997) for a selection of PBC empirical studies covering different nations.

A set of governance issues can be derived from PBC analyses and the three features above:

- Co-ordination of investment decisions and financial commitments with supporting innovation policies.
- Resolution of income distribution conflicts arising out of structural transition through technological change by the use of prices and incomes policies.
- Broadening the base of participation to all citizens through indicative planning processes.
- New assignment of policy instruments over local, regional, national and international jurisdictions; with closer interaction across jurisdictions.
- Alter the electoral period to allow for longer electoral cycles that can produce more stable and longer-term political perspectives.
- Depoliticisation of some elements of the policy agenda by independent sources of economic decision-making (e.g. Central



Bank independence, non-politicised public servants).

The research on PBCs needs to broaden its agenda to incorporate a framework of analyses that includes more of the factors identified in all the various versions of the PBC, and then crucial governance issues should be investigated using this framework. Only then will the PBC concept serve a very useful purpose in the field of public policy.

### Internet Sites

Economic History Services. *Encyclopedia*.  
[www.eh.net/encyclopedia/heckelman.political.business.cycles.php](http://www.eh.net/encyclopedia/heckelman.political.business.cycles.php)

Centre for Economic Policy Research.  
Macroeconomic Policies: Political Business Cycles.  
[www.cepr.org/pubs/bulletin/dps/dp608.htm](http://www.cepr.org/pubs/bulletin/dps/dp608.htm)

### Selected References

Alesina, A. (1989) "Politics and Business Cycles in Industrial Democracies", *Economic Policy*, Volume 4, Number 1, pp. 55-98.

Aschauer, D.A. (2000) "Do States Optimize? Public Capital and Economic Growth", *Annals of Regional Science*, Volume 34, Number 3, pp. 343-63.

Bhaduri, A. and J. Steindl. (1983) "The Rise of Monetarism as a Social Doctrine", *Thames Papers in Political Economy*, Autumn, pp. 1-18.

Boddy, R. and J. Crotty. (1975) "Class Conflict and Macro-Policy: The Political Business Cycle", *Review of Radical Political Economics*, Volume 7, Number 1, pp. 1-19.

Catley, R. and B.J. McFarlane. (1981) *Australian Capitalism in Boom and Depression*. Chippendale, Sydney: Alternative Publishing Cooperative.

Courvisanos, J. (1996) *Investment Cycles in Capitalist Economies: A Kaleckian*

*Behavioural Contribution*. Cheltenham, UK: Edward Elgar.

Franzese, R.J. (2002) "Electoral and Partisan Cycles in Economic Policies and Outcomes", *Annual Review of Political Science*, Volume 5, pp. 369-421.

Frey, B.S. (1978) "Politico-Economic Models and Cycles", *Journal of Public Economics*, Volume 9, Number 2, pp. 203-20.

Frey, B.S. (1997) *Political Business Cycles*. Cheltenham, UK: Edward Elgar.

Galbraith, J.K. and W. Darity, Jr. (1994) *Macroeconomics*. Boston: Houghton Mifflin.

Hallerberg, Mark and Lúcio de Souza. (2000) *The Political Business Cycles of EU Accession Countries*. Tinbergen Institute Discussion Paper 2000-085/2. Amsterdam: Erasmus University.  
[www.tinbergen.nl/discussionpapers/00085.pdf](http://www.tinbergen.nl/discussionpapers/00085.pdf)

Hibbs, D.A. (1977) "Political Parties and Macroeconomic Policy", *The American Political Science Review*, Volume 71, pp. 1467-87.

Kalecki, M. (1943) "Political Aspects of Full Employment", *Political Quarterly*, Volume 14, Number 4, pp. 322-31. Reprinted in Osiatyński (1990) pp. 347-56. Abbreviated version in Kalecki (1971) pp. 138-45.

Kalecki, M. (1945) "Full Employment by Stimulating Private Investment", *Oxford Economic Papers*, New Series 7, pp. 83-92. Reprinted in J. Osiatyński (1990) (Editor), *Collected Works of Michal Kalecki*, Volume I, Oxford: Oxford University Press, 1990, pp. 377-86.

Kalecki, M. (1971) *Selected Essays on the Dynamics of the Capitalist Economy, 1933-1970*. Cambridge: Cambridge University Press.

- Kennedy, M.C. (1973) "Employment Policy: What Went Wrong?", in J. Robinson (Editor), *After Keynes*. Oxford: Basil Blackwell, pp. 71-87.
- Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*. London: Macmillan.
- Mair, D. and A.J. Laramie. (2002) "Full Employment: Gift Horse or Trojan Horse?", *Review of Social Economy*, Volume 60, Number 4, pp. 567-93.
- Nordhaus, W. (1975) "The Political Business Cycle", *Review of Economic Studies*, Volume XLII, pp. 169-90.
- Osiatyński, J. (1990) *Collected Works of Michał Kalecki. Volume I: Capitalism: Business Cycles and Full Employment*. Oxford: Clarendon Press.
- Robinson, J. (1976) "Michał Kalecki: A Neglected Prophet", *New York Review of Books*, 4 March, p.30.
- Steindl, J. (1979) "Stagnation Theory and Stagnation Policy", *Cambridge Journal of Economics*, Volume 3, Number 1, pp. 1-14.
- Tufte, E.F. (1978) *Political Control of the Economy*. Princeton: Princeton University Press.
- Wright, G. (1974) "Political Economy of New-Deal Spending—Econometric Analysis", *Review of Economics and Statistics*, Volume 56, Number 1, pp. 30-8.
- Zarnowitz, V. (1985) "Recent Work on Business Cycles in Historical Perspective: A Review of Theories and Evidence", *Journal of Economic Literature*, Volume XXIII, June, pp. 523-80.

Jerry Courvisanos  
 Centre for Regional Innovation and  
 Competitiveness, School of Business  
 University of Ballarat, Australia  
 j.courvisanos@ballarat.edu.au

## Productivity

*Gianluca Carnabuci*

### **Introduction**

In its most general meaning, productivity is a ratio of output to input, denoting the rate at which resources are converted into goods and services. This definition is straightforward and universally accepted. However, it is also operationally ambiguous, as is reflected by the plurality of conceptualizations present in the scientific literature.

Productivity is a concept central to virtually all aspects of economic activity. As Krugman (Krugman 1997) puts it, “[it] isn’t everything, but in the long run it is almost everything”. Research efforts in the productivity field include explaining economic growth and technological progress, and assessing populations’ living standards. Given the societal relevance of such issues, it is no surprise that productivity is not only a central and controversial concept in the academic sphere, but also among policy-makers. Institutionalized think tanks on productivity issues have been established almost everywhere in the developing and developed world, to inform policy decisions.

### **Theoretical Frameworks**

It is only since the seminal papers of Nobel Prize laureate Robert Solow (1956, 1957) that the concept of productivity gains full prominence in the scientific debate. The so-called Solow model is, in fact, not a theory of productivity. Rather, it explains the effects of productivity on economic growth. Nonetheless, the model provides the first clear framework for the analysis of productivity. The Solow model revolves

around the Cobb-Douglas production function, thus it features the typical neoclassical assumptions. Mathematically, given an economy’s production function is described by equation (1):

$$(1) \quad Y = AK^{\alpha}L^{(1-\alpha)}$$

where  $0 \leq \alpha \leq 1$ ,  $L$  stands for labor,  $K$  for capital, and  $A$  for technology, then:

$$(2) \quad \frac{\Delta Y}{Y} = \frac{\Delta A}{A} + \alpha \frac{\Delta K}{K} + (1-\alpha) \frac{\Delta L}{L}$$

In (2),  $\frac{\Delta Y}{Y}$ ,  $\alpha \frac{\Delta K}{K}$ , and  $(1-\alpha) \frac{\Delta L}{L}$  represent, respectively, growth in output, capital, and labour, while  $\frac{\Delta A}{A}$  measures productivity growth.

The Solow model achieves two major goals. First, it makes possible to deduce from the production function, effects and restrictions on the productivity of economic systems. Second, it provides a specific methodology for the empirical measurement of productivity. By analyzing long-run national account data, Solow was able to demonstrate that historical output growth largely outpaces the weighted average increase in combined inputs. The large gap between the two growth rates, now known as “Solow residual”, demonstrates that economic growth is not just a matter of increasing inputs, but is to a large extent a matter of how efficiently these inputs are utilized, i.e. a matter of productivity.

Due to its elegance and direct applicability to empirical research, the neoclassical framework proposed by Solow has become the backbone of the productivity field in the academic, policy-making, and business sphere. The model, however, has two limitations. The first is

that the “Solow residual” is not a very accurate measure of productivity. Productivity growth is caused either by technological progress or by efficiency gains, such as the elimination of spurious bureaucracy or the reduction of social conflict. However, the “Solow residual” does not distinguish between these two classes of factors. Instead, by postulating efficiency at all times, the Solow model overestimates the effects of technological progress. In addition, the “Solow residual” is a biased measure of productivity. Precisely because it is a residual, it confounds the real effects of productivity with those of measurement errors, omitted variables, aggregation biases, and model misspecifications, which are unlikely to be randomly distributed. For this reason, Abramovitz (1956) suggested that the “Solow residual” is best understood as a “measure of our ignorance”.

The second limitation of the Solow model is theoretical: productivity is modelled as an exogenous factor. Therefore, the theory is unable to explain where productivity comes from and how it can be encouraged.

No matter how critically one looks at it, the Solow model not only set the stage of the productivity debate, but it still permeates it. Subsequent efforts have tried to elaborate the model in order to overcome its principal limitations. These attempts have bifurcated in two main lines.

A first line of research has concentrated on developing refined versions of the Solow model. These are commonly referred to as neoclassical growth models, reflecting their common effort to leave the neoclassical assumptions of the Solow model untouched. Based on increasingly fine-grained empirical specifications and growth accounting techniques, the primary

goal of these models is to enhance the quality of the Solow residual as a measure of productivity. The basic idea is that the more precisely the accumulated inputs that contribute to production are accounted for, the more precisely the residual will measure the effects of productivity. Neoclassical growth models have developed in multiple directions. Notably, Jorgenson and Griliches (1967) take into account the heterogeneity of capital by creating a constant quality index. In the same spirit, Griliches (1960), Denison (1962), and Jorgenson and Griliches (1967) suggest that labor cannot be simply modeled as a homogeneous factor, and distinguish between quantity and quality of labor. Elaborating on this idea, the so-called augmented Solow models (Mankiw et al 1992) introduce human capital as a production factor in its own right. Finally, in an attempt to account for the role of knowledge in modern economies, Griliches (1973, 1979) includes R&D as yet another production factor.

By reducing the “measure of our ignorance”, the neoclassical growth models have greatly contributed to our understanding of productivity. Their usefulness is widely recognized also outside the academic sphere, where bodies of the U.S. Government, such as the Social Security Administration, the Congressional Budget Office, the Office of Management and Budget, and the General Accounting Office, use neoclassical growth models for their long-run projections (Stiroh 1998b). However, because they adopt the same neoclassical theoretical framework as Solow does, they cannot provide a satisfactory explanation as to what drives productivity growth.

A second line of research, which goes under the name of New Growth Theory or

endogenous growth models, departs from the purely neoclassical approach. The New Growth Theory originates from a fundamental theoretical critic of the Solow model: productivity is not an exogenous factor in economic dynamics. To the contrary, it changes as the result of actions taken by economic agents, which in turn depend on the conditions that the Solow model tries to explain. Therefore, productivity should be explicitly modeled as an endogenous variable that at the same time explains and is explained by the fundamental dynamics of economic production. Arrow (1962) had already suggested that productivity could display an endogenous dynamic, due to the productivity-enhancing effects of “learning by doing”, which he modeled as a function of cumulative capital stock. However, it is Romer (1986) who formalized the first full-fledged endogenous growth model. Building on Arrow’s intuition, Romer postulates that productivity is primarily driven by knowledge, which has the characteristics of a public good: once it is produced locally, it tends to spill over to the entire economy. Accordingly, “[t]he creation of new knowledge by one firm is assumed to have a positive external effect on the production possibilities of other firms” (Romer 1986:1003). This induces a self-reinforcing cycle of knowledge externalities, which, by offsetting at the aggregate level the diminishing returns from other inputs, sustains long-term productivity growth. Romer’s conclusion was a breakthrough. Knowledge was not anymore just a factor of production, as a few neoclassical models had already acknowledged, but *the* engine of productivity. As such, Romer’s theory is what Solow’s was not: a theory of productivity. If one wants to understand

what drives productivity growth, one needs to investigate the complex functioning of knowledge-enhancing factors such as education (e.g., Lucas 1988), on-the-job training (Bartel 1995), research and development (R&D) (e.g., Romer 1990), and academic production (Adams 1990).

### Measures of Productivity

There exist a variety of empirical measures of productivity, and the choice of which one to use ultimately depends on the purpose of the research, as well as on theoretical, methodological, and data-availability issues. Perhaps the most fundamental choice is between measures of total factor productivity (TFP) and measures of single-factor productivity. TFP measures quantify the “Solow residual”, i.e. the gap between output growth and the weighted growth of input factors, expressed by  $\frac{\Delta A}{A}$  in equation (2). Because

of the unfeasibility to measure empirically the growth of all input factors, TFP is now also commonly referred to as multi-factor productivity. Typically, TFP studies take into account the growth trajectories of the two fundamental input factors, labour and capital. However, many studies model also the growth changes of important intermediate inputs such as energy, materials, and services. It goes without saying that the more precisely all relevant factors are accounted for in the measure of TFP, the more precisely the researcher will be able to quantify the “pure” effects of productivity. At the same time, however, the more input factors are incorporated in the model, the more methodological problems arise. Next to the obvious problem of data availability, TFP models are bound to the awkward issue of grouping together input factors of

heterogeneous nature. Typically, the solution is to use their monetary values. However, this approach raises another problem, as monetary changes do not necessarily correspond to real changes in production; hence, they may misrepresent productivity dynamics. Unfortunately, while estimations are known to be sensitive to the choice of methodology, there is not an agreed upon solution among researchers on how to best approach this problem. Therefore, although optimal from a theoretical point of view, TFP measures are seen by some critics as lacking sufficient robustness (Sargent & Rodriguez 2000), especially in comparative international analyses.

Measures of single-factor productivity calculate the difference between output growth and the growth of only one specific input factor. Unlike TFP, the concept of single-factor productivity corresponds only imprecisely to the theoretical definition of productivity. In fact, it accounts not only for the productivity of the focal input factor, but also for the accumulation of the other factor(s). However, either for their methodological simplicity, or because they match more closely specific research objectives, measures of single-factor productivity are often preferred to TFP measures. Because it relates directly to people's living standards and it is easily measurable, labour productivity, i.e. output per unit of labour input, is by far the most widely used measure of single-factor productivity. Given (1), this is calculated as in (3):

$$(3) \quad \left( \frac{\Delta(Y/L)}{Y/L} \right) = \frac{\Delta A}{A} + \alpha \left( \frac{\Delta(K/L)}{K/L} \right)$$

In words, growth of labor productivity equals the sum of weighted growth of

capital per worker, plus growth of total factor productivity.

Typically, labour input is empirically quantified either as the average annual number of employees, or as the total number of hours worked in a year. When available, the latter specification is preferable because it makes historical and international comparability possible. The larger amount of hours worked in a year by the average US employee compared to the average European employee, for example, would produce an overestimation of US labour productivity if the number of employees was used as proxy for labour input.

### **International Productivity Trends**

The empirical literature on productivity, especially concerning developed economies, is huge. Let us look at some stylised facts in a historical and international perspective.

Immediately after WWII, the United States was by far the most productive country in virtually every industry. In 1950, for example, labor productivity in the United States was three times higher than in Germany, Austria and Italy, and seven times higher than in Japan (Smolny 2000). Between 1950 and 1973, however, developed countries greatly narrowed their gaps with the United States (Abramovitz 1994). Particularly in the 1950s and 1960s, they lived a “golden age” of productivity growth (Maddison 1995). Somewhat unexpectedly, however, in 1973 productivity growth began to suffer from a generalized slowdown among all developed countries. For the OECD as a whole, it decreased from a rate of nearly 3% per year for the period 1960–73, to only 0.5% between 1973 and 1979 (OECD 1996). In the second half of the 1990s, US

productivity has again begun to grow faster, with an estimated TFP growth for the 1995-98 period nearly three times as high as during the 1973-95 period (Stiroh 2001). Europe, however, has not yet proven able to keep up with this US upsurge.

While in the 1970s Western developed countries were struggling with slower productivity rates, the East Asian economies of Taiwan, Hong Kong, South Korea, and Singapore showed robust economic growth. Thailand, Malaysia, Indonesia, the Philippines, and China then followed in the course of 1980s. The economic growth of the “Asian tigers” in that period, however, can only be partly attributed to an increase in productivity. In fact, as Krugman (1994) and Young (1995) point out, their growth is to a large extent due to an unprecedented mobilization of resources. Nonetheless, as other authors emphasize (Nehru & Dharehwar 1994), in those years productivity growth in East Asia has been remarkably high, and certainly higher than in most other developing regions in the world.

Recently productivity levels have tended to converge among developed countries, and, to a lesser degree, between developing and developed countries. However, the least developed countries are lagging behind. In fact, “[i]n 1988, output per worker was 48 times higher in the most productive compared to the least productive country” (Hall and Jones 1996). There is evidence that the productivity of the least developed countries has decreased on average in the years between 1970 and 1992 (Forstner et al 2001).

### **Sources of Productivity**

There is a large body of empirical research on the factors driving productivity.

Productivity appears to be highly correlated with investment in physical capital, and the correlation has proven robust across long historical periods, and in both developed and developing countries. In particular, investment in equipment and machinery has the strongest effects, while the results for non-equipment investments tend to be weaker. The effects of investment in physical capital, moreover, seem to be particularly strong in developing countries in transition between a traditional “Malthusian” economy and an industrialized one (Landon-Lane & Robertson 2003).

In 1987, Solow unveiled what is often referred to as the “productivity paradox”: “You can see the computer age everywhere but in the productivity statistics” (Solow 1987). Ever since, a great deal of attention has been paid to assessing whether the much-praised IT revolution was in fact having any significant impact on productivity growth. The extraordinary upsurge in US productivity of the late 1990s, however, has convinced most economists that investments in IT have at last begun to bear fruits. Moreover, today many researchers realize that the productivity statistics Solow was referring could not capture much of the benefits computers bring in the service sectors (McGuckin & Stiroh 2002).

If nowadays there is large consensus that the diffusion of IT has been enhancing productivity growth, the debate is still lively when it comes to assessing the nature of IT-driven effects. On the one hand, optimists find evidence that information technologies have sparked a revolution comparable to those witnessed when electricity or the internal combustion engine diffused within the system of economic production. Hence, they argue



that the recent productivity resurgence reflects a permanent shift in the performance of US economy (Jorgenson & Stiroh 2000; Oliner & Sichel 2000; Whelan 2002).

On the other hand, skeptics claim that far too much uncertainty remains to draw such buoyant conclusion. In an influential article, Gordon (2000), for example, shows that much of the observed increase in US productivity could be a spurious effect of the underlying economic cycle, rather than an effect of IT. Moreover, he finds that, unlike technologies that brought about permanent shifts in productivity in the past, the productivity-enhancing effects of IT have not significantly impacted on the economic system at large, but have remained confined to the IT-producing sectors (Gordon 1999). As urgent as it may be, a definitive answer to the question of whether U.S. economy has entered a new era of long-term productivity growth is not possible at this stage.

Many studies suggest that openness to international trade is conducive to productivity growth. While much of the empirical research on the subject is affected by methodological limitations (see Edwards 1993), Edwards (1998) shows the positive and significant relationship between openness and TFP growth to be robust to different indicators, estimation techniques, time periods, and functional forms, both in developed and developing countries.

Human capital, particularly in the form of education, is shown to be another effective driving force of productivity growth (Barro & Sala-i-Martin 1999; Levine & Renelt 1992; Nehru & Dharehwar 1994). Its impact varies greatly according to the level of a country's development, as less developed countries

benefit most from it (Benhabib & Spiegel 1994). Moreover, while primary and secondary education appear to increase productivity in underdeveloped and developing economies respectively, in OECD countries productivity grows mostly with advanced education (Gemmell 1996). In addition to its direct contribution, there is substantial evidence that human capital contributes indirectly to productivity by stimulating the accumulation of physical capital and technology, and by bettering health, while discouraging population growth and infant mortality (Doyle & Weale 1994; Wolff & Gittleman 1993). Finally, human capital seems to increase productivity by favouring the formation of a "social infrastructure" conducive to economic production, i.e. institutions able to effectively protect returns to investment from predation, and to favour linguistic and economic international openness (Hall & Jones 1999).

Much research has been done to investigate the contribution of investment in R&D to productivity growth, and there is compelling evidence that R&D fosters technological advancement and, thus, productivity. There are different types of R&D, and their effects seem to work through different channels. A major difference is between public and business R&D. While public R&D is thought to contribute to long-term productivity mainly by enlarging the stock of scientific knowledge available to society, this argument cannot be easily tested empirically. In one of the very few studies on the subject, Adams (1990) finds that the stock of fundamental knowledge, as measured by accumulated academic scientific papers, significantly relates to productivity growth in US manufacturing industries. Unlike public R&D, business



R&D has been investigated in many empirical studies, at all levels of aggregation, and for many countries. The evidence is compelling that business R&D enhances productivity by innovating products and production processes (see Nadiri (1993) for a survey of this literature). Another distinction should be drawn between domestic and foreign R&D, as the latter is found to be particularly important for OECD countries, with the notable exception of the United States. At any rate, countries that invest in their own R&D appear to benefit most from foreign R&D (Guellec & van Pottelsberghe 2001).

### **Policy Implications**

Unfortunately, so much empirical research has not yet been sufficient to settle the theoretical controversy that divides neoclassical and new growth theorists, as productivity data, typically, are too coarse to discern specific causal mechanisms. The different interpretations of the empirical evidence, advanced by neoclassical and new growth theorists, are not just an academic dispute. To the contrary, they lead to profoundly different policy implications. Neoclassical growth theorists maintain that long-run productivity depends exclusively on technological progress, while all other factors may have a beneficial effect only in the short and medium term. Moreover, as we have seen, they do not address the problem of how technological progress can be encouraged. Accordingly, they suggest no specific policy interventions with regard to long-term productivity, and praise the role of market allocation instead. New growth theorists, on the contrary, believe that long-term productivity is sustained by the positive externalities associated with the creation of knowledge. Since knowledge is

a mixed good, they conclude that pure market dynamics lead to socially inefficient allocations. As a consequence, while they concord with the neoclassical idea that technological progress is an essential driving force of long-term productivity, they believe that policy interventions are necessary whenever knowledge spillovers are generated.

New growth theorists in the tradition of Arrow (1962), for example, maintain that investment in equipment generates effects of “learning by doing” that tend to spill over to other firms (De Long & Summers 1991, 1992, 1994). Therefore, they posit that investments in physical capital should be encouraged by government policies because that, in turn, will increase long-term productivity. In this vein, De Long and Summers (1992) suggest that governments should minimize ‘anti-equipment’ incentives by ensuring the effective protection of property rights and a stable macro economic. In fact, most OECD countries have made significant steps towards stability-oriented macroeconomic policies (Sanghoon & Hemmings 2000), and property right systems are being increasingly harmonized across national borders. More specific pro-equipment policy measures adopted by governments, moreover, are investment grants, tax relief on investment, and infrastructure expenditure.

Some researchers claim that public institutions can directly enhance the productivity of economic systems by public investments in infrastructure. Empirical evidence on the productivity of public spending, however, is mixed, and virtually all studies are contested on methodological grounds. So, while Aschauer (1989) and Munnell (1990) find public capital to increase significantly

productivity, Evans and Karras (1994) and Holtz-Eakin (1994) conclude that no such effects exist.

New growth theorists maintain that also economic openness is conducive to long-term productivity growth, because it channels bits of the international knowledge stock towards local economies, via imitation and use (Grossman & Helpman 1991; Obstfeld & Rogoff 1996). Because of the advanced technologies it gives access to, in fact, new growth theorists regard policies of trade liberalization as a precondition for developing countries to catch up with developed ones. However, especially since the country studies carried out by Bhagwati (1978), and Krueger (1977), they recognize that a policy of trade liberalization is by itself insufficient, and must be complemented by a more comprehensive outward-looking policy. This should include a stable and non-discriminatory exchange-rate system, prudent monetary and fiscal policies, and corruption-free policy administrations (Baldwin 2003).

Human capital is another form of knowledge that, according to the new growth theorists, is not fully internalised by who possesses it; rather, it tends to spill over to other employees, firms, industries, and even nations. Therefore, in their view, policies should be designed that stimulate people to invest in education, firms to pay higher salaries to educated workers and invest in post-educational training, and that improve the quality of educational institutions. Another instrument that is sometimes suggested to enhance an economy's human capital is the immigration of skilled labour.

A central theme of the new growth theory is R&D, which is regarded as a major source of knowledge externalities

and, therefore, as an issue of primary importance in long-term productivity policies. Public policy can intervene in two ways. First, it can stimulate investment in R&D. Interventions to this end include direct government funding of private R&D; tax incentives for private R&D; public-sector R&D (including academic research); provisions for the diffusion of research; and, improvement of the legal protection of intellectual property rights. Second, it can stimulate spillover effects. Policy instruments to this end include the promotion of technology parks; the agglomeration of key decision-making units such as head offices; the development of technological standards; the improvement of specialized financial market mechanisms (e.g. venture capital); and, cooperative R&D projects. There are many different structures of R&D financing. Although the business sector is overall the predominant source, non-business investment is substantial even in "free market" oriented economies.

The OECD (1996a) demonstrate that, in developed countries, the most widely used financial instruments to stimulate R&D are research grants and tax concessions. Moreover, the data seem to indicate a tendency of developed countries to intervene more by direct funding of research grants, than by tax concessions. This is interesting, as direct funding is generally believed to favour fundamental over commercially-oriented research. In developing countries, R&D policies are often constrained by the governments' need to respond to the political demand of increasing market liberalization. This is especially true for policies based on direct public funding. However, some of the most successful innovators, such as Korea,

Taiwan and Singapore, do feature an articulated R&D public policy.

### Future Research

The productivity field has advanced greatly in the last fifty years. However, the theoretical controversy between neoclassical and new growth theory has not yet been conclusively settled. This is the most crucial issue in productivity research, especially in view of the widely recognized need of scientifically informed public policies.

Although much research has already been done to test the two theories, recently collected TFP, panel, and firm-level data lend themselves to more specific hypothesis testing, and should be fully investigated to this purpose. Notwithstanding the larger amount of fine-grained data currently available, moreover, there are still a number of important deficiencies. Of these, perhaps the most crucial is the lack of estimates of knowledge capital, which should include information on patent stocks, R&D expenditures, and employer education and training. Finally, given that much of the dispute between neoclassical and new growth theorists concerns the functioning of systems of knowledge production, productivity research should explicitly address this issue. Since standard economic modelling is cumbersome to this purpose, theoretical and analytical approaches coming from outside the realm of economics should be welcomed. Research on national systems of innovation (NSI), and social network models of knowledge and technology diffusion, are prime examples in this direction.

### Internet Sites

National Bureau of Economic Research.  
[www.nber.org/programs/pr/pr.html](http://www.nber.org/programs/pr/pr.html)  
OECD [www.oecd.org](http://www.oecd.org)  
US Bureau of Labor statistics.  
[www.bls.gov/home.htm](http://www.bls.gov/home.htm)

### Selected References

- Abramovitz, M. (1994) "The Origins of the Postwar Catch-up and Convergence Boom", in N. von Tunzelmann (Editor), *The Dynamics of Technology, Trade and Growth*. Aldershot, UK: Edward Elgar, pp. 21-52.
- Abramovitz, Moses. (1956) "Resource and Output Trends in the United States Since 1870", *American Economic Review*, Volume 46, pp. 5-23.
- Adams, James D. (1990) "Fundamental Stocks of Knowledge and Productivity Growth", *Journal of Political Economy* Volume 98, pp. 673-702.
- Arrow, Kenneth J. (1962) "The Economic Implications of Learning by Doing", *Review of Economic Studies*, Volume 29, pp. 155-173.
- Aschauer, D.A. (1989) "Is Public-Expenditure Productive", *Journal of Monetary Economics*, Volume 23, pp. 177-200.
- Baldwin, R. E. (2003) "Openness and Growth: What's the Empirical Relationship?" *NBER Working Paper*. New York: NBER.
- Barro, Robert J. and Xavier Sala-i-Martin. (1999) *Economic Growth*. Cambridge, Mass.: MIT Press.
- Bartel, A. (1995) "Training, Wage Growth and Job Performance: Evidence from a Company Database", *Journal of Labor Economics*, Volume 13, pp. 401-425.
- Benhabib, J. and M.M. Spiegel. (1994) "The Role of Human-Capital in Economic-Development Evidence from Aggregat Cross-Country Data", *Journal*

- of *Monetary Economics*, Volume 34, pp. 143-173.
- Bhagwati, J. (1978) *Foreign Trade Regimes and Economic Development: Anatomy and Consequences of Exchange Control Regimes*. New York: National Bureau of Economic Research.
- De Long, J B and L H Summers. (1991) "Equipment Investment and Economic Growth", *The Quarterly Journal of Economics*, Volume 106, pp. 445-502.
- De Long, J B and L H Summers. (1992) "Equipment Investment and Economic Growth: How Strong is the Nexus?" *Brookings Papers on Economic Activity*, Number 2.
- De Long, J B and L H Summers. (1994) "Equipment Investment and Economic Growth: Reply", *Quarterly Journal of Economics*, pp. 803-807.
- Denison, Edward Fulton. (1962) *The Sources of Economic Growth in the United States and the Alternatives Before Us*. New York: Committee for Economic Development.
- Doyle, C. and M. Weale. (1994) "Education, Externalities, Fertility and Economic Growth." *Education Economics*, Volume 2, pp. 129-167.
- Edwards, S. (1993) "Openness, Trade Liberalization, and Growth in Developing-Countries", *Journal of Economic Literature*, Volume 31, pp. 1358-1393.
- Edwards, S. (1998) "Openness, Productivity and Growth: What Do We Really Know?", *Economic Journal*, Volume 108, pp. 383-398.
- Evans, P. and G. Karras. (1994) "Is Government Capital Productive—Evidence from a Panel of 7 Countries", *Journal of Macroeconomics*, Volume 16, pp. 271-279.
- Forstner, Helmut, Anders Isaksson, and Thiam Hee Ng. (2001) "Growth in Least Developed Countries: An Empirical Analysis of Productivity Change, 1970-1992", *Sin Working Paper Series*.
- Gemmel, N. (1996) "Evaluating the Impacts of Human Capital Stocks and Accumulation on Economic Growth: Some New Evidence", *Oxford Bulletin of Economics and Statistics*, Volume 58, pp. 9-28.
- Gordon, R. (2000) "Does the 'New Economy' Measure Up to the Great Inventions of the Past?" *Journal of Economic Perspectives*, Volume 14, pp. 49-74.
- Griliches, Z. (1973) "Research Expenditures and Growth Accounting in Science and Technology", in B.R. Williams (Editor), *Economic Growth*. London: Macmillan, pp. 59-95.
- Griliches, Z. (1979) "Issues in Assessing the Contribution of Research and Development to Productivity Growth", *Bell Journal of Economics*, Volume 10, pp. 92-116.
- Griliches, Zvi. (1960) "Measuring Inputs in Agriculture: A Critical Survey", *Journal of Farm Economics*, Volume 42, pp. 1411-33.
- Grossman, Gene M. and Elhanan Helpman. (1991) *Innovation and Growth in the Global Economy*. Cambridge, Mass.: MIT Press.
- Guellec, D and Bruno van Pottelsberghe. (2001) "R&D and Productivity Growth: Panel Data Analysis of 16 OECD Countries", *STI Working Paper*.
- Hall, R. E. and C. I. Jones. (1999) "Why Do Some Countries Produce So Much More Output Per Worker than Others?", *Quarterly Journal of Economics*, Volume 114, pp. 83-116.

- Hall, Robert E. and Charles I. Jones. (1996) "The Productivity of Nations." *NBER Working Paper 5812*. New York: NBER.
- Holtz-Eakin, D. (1994) "Public-Sector Capital and the Productivity Puzzle", *Review of Economics and Statistics*, Volume 74, pp. 12-21.
- Jorgenson, D. W. and K. J. Stiroh. (2000) "Raising the Speed Limit: US Economic Growth in the Information Age", *Brookings Papers on Economic Activity*, pp. 125-235.
- Jorgenson, Dale W. and Zvi Griliches. (1967) "The Explanation of Productivity Change", *Review of Economic Studies*, Volume 34, pp. 249-280.
- Krueger, A. (1977) "Trade Policy and Economic Development: How We Learn", *American Economic Review*, Volume 87, pp. 1-22.
- Krugman, P. (1994) "The Myth of Asia's Miracle", *Foreign Affairs*, Volume 73, pp. 62-78.
- Krugman, Paul R. (1997) *The Age of Diminished Expectations U.S. Economic Policy in the 1990s*. Cambridge, Mass.: The MIT Press.
- Landon-Lane, J. and P. Robertson. (2003) "Accumulation and Productivity Growth in Industrializing Economies." Rutgers University Working Papers.
- Levine, R. and D. Renelt. (1992) "A Sensitivity Analysis of Cross-Country Growth Regressions", *American Economic Review*, Volume 82, pp. 942-963.
- Maddison, Angus. (1995) *Monitoring the World Economy, 1820-1992*. Paris: Development Centre of the Organisation for Economic Co-operation and Development.
- Mankiw, N. Gregory David Romer, and David N. Weil. (1992) "A Contribution to the Empirics of Economic Growth", *Quarterly Journal of Economics*, Volume 107, pp. 407-37.
- McGuckin, R. and K. Stiroh. (2002) "Do Computers Make Output Harder to Measure?", *Economic Inquiry*, Volume 40, pp. 42-59.
- Munnell, A. (1990) "Why Has Productivity Growth Declined? Productivity Growth and Public Investment", *New England Economic Review*, pp. 3-22.
- Nadiri, I. (1993) "Innovations and Technological Spillovers", *NBER Working Paper Series*. New York: NBER.
- Nehru, V. and A. Dhareshwar. (1994) *New Estimates of Total Factor Productivity Growth for Developing and Industrial Countries*. Washington D.C.: The World Bank.
- Obstfeld, M. and K. Rogoff. (1996) *Foundations of International Macroeconomics*. Cambridge, MA: MIT Press.
- OECD. (1996) *Technology, Productivity and Job Creation*. Paris: OECD.
- Oliner, S. D. and D. E. Sichel. (2000) "The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?", *Journal of Economic Perspectives*, Volume 14, pp. 3-22.
- Romer, Paul. (1986) "Increasing Returns and Long-Run Growth", *Journal of Political Economy*, Volume 94, pp. 1002-1037.
- Sanghoon, A. and P. Hemmings. (2000) "Policy Influences on Economic Growth in OECD Countries: An Evaluation of the Evidence", *OECD Economic Working Paper*. Paris: OECD.
- Sargent, T. and E. Rodriguez. (2000) "Labor or Total Factor Productivity: Do We Need to Choose?", *International*

- Productivity Monitor*, Volume 1, pp. 41-44.
- Setterfield, M. and J. Cornwall. (2002) "A Neo-Kaldorian Perspective on the Rise and Decline of the Golden Age", in M. Setterfield (Editor), *The Economics of Demand-Led Growth*. Cheltenham, UK: Edward Elgar.
- Smolny, W. (2000) "Post-War Growth, Productivity Convergence and Reconstruction", *Oxford Bulletin of Economics and Statistics*, Volume 62, pp 589-606.
- Solow, Robert M. (1987) "We'd Better Watch Out", *New York Times Book Review* (July 12): 36.
- Solow, Robert M. (1956) "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics*, Volume 70, pp. 65-94.
- Solow, Robert M. (1957) "Technical Change and the Aggregate Production Function", *Review of Economics and Statistics*, Volume 39, pp. 312-320.
- Stiroh, K. 2001. "What Drives Productivity growth?" *Economic Policy Review*, pp. 37-59.
- Stiroh, K. J. (1998b) "Long-run Growth Projections and the Aggregate Production Function: A Survey of Models used by the U.S. Government", *Contemporary Economic Policy*, Volume XVI, pp. 467-479.
- Weisskopf, T , S Bowles, and D Gordon. (1983) "Hearts and Minds: A Social Model of US Productivity Growth", *Brookings Papers on Economic Activity*, pp. 381-441.
- Whelan, K. (2002) "Computers, Obsolescence, and Productivity", *Review of Economics and Statistics*, Volume 84, pp. 445-461.
- Wolff, E N , and M Gittleman. (1993) "The Role of Education in Productivity Convergence: Does Higher Education Matter?", in B van Ark and D Pilat (Editors), *Explaining Economic Growth*. Amsterdam: North Holland.
- Young, A. (1995) "The Tyranny of Numbers—Confronting the Statistical Realities of the East-Asian Growth Experience", *Quarterly Journal of Economics*, Volume 110, pp. 641-680.

Gianluca Carnabuci  
 Institute of Management and Organization  
 University of Lugano  
 CH-6904 Lugano, Switzerland  
 gianluca.carnabuci@lu.unisi.ch

## Public Goods, Externalities and Governance

*Jouni Paavola*

### **Introduction**

Externalities and public goods—together with monopolies, economies of scale and other deviations from the assumptions of the competitive model—are cornerstones of “the market failure” paradigm. It holds that when markets cannot deliver efficient outcomes on their own—such as when externalities or public goods are present—the government has to establish and implement public policies calculated to achieve optimal allocation of resources and to maximise social welfare (Cowen 1988; Zerbe and McCurdy 1999).

The beginnings of the market failure paradigm can be traced to Pigou’s (1929) seminal work in welfare economics. He identified divergences between private and social net marginal products as obstacles for maximising social welfare and suggested that government intervention could avoid sub-optimal outcomes. Paul Samuelson (1954), Richard Musgrave (1959) and others consolidated the market failure paradigm in the 1950s. However, the market failure paradigm was contested since its inception by, for example, Knight (1924), Tiebout (1956), Coase (1960, 1974) and Buchanan (1965). The critics have shown that the market failure paradigm’s arguments regarding market failures and institutional responses to them have often been inconsistent or unfounded.

The market failure paradigm has a narrow, state-centred view of the range of institutional solutions that can be used as a response to market failures. This view deviates significantly from the contemporary political reality. States are currently

relinquishing some of their governance authority to emerging regional and international entities such as the European Union, NAFTA, WTO and the United Nations. States have also delegated governance authority to sub-national and local governments and have facilitated the participation of firms, non-governmental organisations and other stakeholders in collective decision-making. States have also recognised the authority of some stakeholders to self-governance.

This transition to the use of a greater range of institutional solutions is sometimes called as one from “government” to “governance”. However, this is misleading as “the government” is still intimately involved in governance. The government has broadened the range of governance solutions as a response to the emergence of more complex and far-reaching governance problems and increasingly pluralist governance goals. Moreover, the government is needed to back up the authority and legitimacy of new governance solutions.

While it is fair to say that there is no comprehensive economic theory of governance yet, it is obvious that it will be based in part on the scholarship on public goods and externalities that has underpinned the market failure paradigm. A reinterpretation of externalities, public goods and other market failures in the light of their criticisms can offer a nuanced understanding of governance problems and institutional responses to them. To put it differently, while it was enough for the market failure paradigm to call for government intervention when an instance of a market failure was observed, the emerging governance paradigm must be able to characterise governance problems in greater detail, to identify a range of applicable governance



solutions, and to assess their likely consequences and desirability in pluralist instead of only welfarist light.

In what follows, the second section examines the notions of public goods and their criticisms in greater detail. The third section reviews the notions of externalities and their criticisms. The fourth section discusses how the scholarship on public goods and externalities can contribute to a theory of governance.

### **Definitions and Analysis of Public Goods**

Central and local governments and various organisations have provided public goods such as religious rites, public safety and defence from the time immemorial. While John Stuart Mill and Henry Sidgwick identified the challenge of supplying them already in the 19th century, public goods did not receive close attention in economics before the mid-20th century. In his ground-breaking contribution, Samuelson (1954: 387) defined public goods as “collective consumption goods ... which all enjoy in common in the sense that each individual’s consumption of such a good leads to no subtraction from any other individual’s consumption of that good”. Others have later called this attribute of certain goods as *joint consumption* (Head 1962). Samuelson (1954) concluded that the market system does not make available an optimal amount of jointly consumed goods and that for this reason they should be publicly provided.

Samuelson’s views of public goods did not remain uncontested. Some, such as Margolis (1955), criticised his definition of public goods as not matching with the attributes of many publicly provided goods. Others contested the idea that public goods are enjoyed by all by arguing that it is possible to exclude users from many

publicly provided goods and that this is actually done with many of these goods.

One significant confusion behind these criticisms was the misidentification of public goods with public provision and non-exclusive access. It has not been uncommon for governments to provide private goods to their citizens but their having done so does not transform these goods into “public goods”. For example, Musgrave (1959) argued that there are *merit goods* such as education and health care that people would consume too little if their supply was left to the markets. Merit goods are essentially private goods but it is thought that people’s preferences regarding them are not well-informed because of imperfect information, irrationality and other reasons (Head, 1974). Conversely—and was discussed by Buchanan (1965) and Coase (1974)—public goods can sometimes be provided privately without public involvement.

This confusion indicates that the hallmark of public goods was not clearly identified. Joint consumption should be considered the most important attribute of public goods. It has important implications for pricing and recovering the costs of public goods. Samuelson demonstrated that any positive price would render the consumption of public goods sub-optimal (1958:335). Yet some, such as Goldin (1977), downplayed the possibility of joint consumption because they saw it as an undiminished consumption by an infinite number of users. As that is indeed hardly conceivable, for many the choice appeared to be simply between exclusive and open access to goods.

Buchanan (1965) usefully pointed out that the scale of joint consumption can vary as the result of indivisibility or “lumpiness” of some goods. For instance, some goods and services such as live puppet shows can be jointly consumed by only a small number



of agents, while others such as the rule of law can serve a great number of them. *Crowding* was also considered to characterise certain public goods. However, all public goods can become crowded when their capacity to provide for joint consumption is surpassed.

The implications of the scale of joint consumption were first brought up by Tiebout (1956), who argued that Samuelson's conclusions and treatment of public goods were based on a simplistic view according to which the central government provides all public goods which are consumed jointly by all citizens. Tiebout argued that Samuelson had ignored the role of local governments in the provision of many public goods such as fire services, parks, recreational facilities and local amenities. He sought to refute Samuelson's conclusion according to which no decentralised price system can generate optimal level of public goods consumption by demonstrating how the provision of public goods by local governments would result in an optimal outcome when the consumers are perfectly mobile and there are an infinite number of local jurisdictions. The consumers would choose to reside in a jurisdiction which provides the bundle of public goods and the level of indirect prices (tax burden) best matching their preferences (see also Buchanan 1972).

The relationships between public goods, excludability, and *de facto* exclusion have been another subject of long-lasting ambiguity. As Goldin (1977) suggested, it is indeed often possible to choose between exclusive and open access to goods, whether they are public or private. However, this is not to say that exclusion is trivial with all goods. As Head (1962) pointed out, joint consumption and excludability are different attributes that may or may not characterise

goods simultaneously. That is, it is easy to exclude users from some jointly consumed goods such as football matches, roads and bridges, while other goods, for example public safety and defence, make exclusion more difficult. It is also noteworthy that excludability is not a dichotomous attribute: the costs of exclusion vary from low to high (Demsetz 1964; Schmid 1987). The level of exclusion costs depends on the size, divisibility, boundaries and other attributes of goods, technologies that are available for the exclusion of unauthorised users, and the attributes of involved agents.

The level of exclusion costs influences the way in which public goods can be provided. When exclusion is costly, the only way to provide a public good may be, as Samuelson (1954) argued, for the government to tax the citizens and to use the tax revenue to provide the public good on a non-exclusive basis. When exclusion is easy, central and local governments can impose user fees such as those for national parks or communal swimming pools. They can also limit access to public goods by some criterion such as merit or need. Many publicly provided goods—whether they are private or public goods—are in fact provided on an exclusive basis.

When exclusion is relatively easy, public goods can also be supplied privately. Coase's analysis of lighthouses—the epitome of public goods for welfare economics—is a case in point. Coase (1974:366-367) demonstrated how in England and Wales lighthouses were predominantly privately built and operated before the end of the 18th century, and how their public building and operation was an explicit policy choice made in the 1830s. Private supply of lighthouses was possible because tolls for their services could be collected from vessels upon their arrival to

harbours. The private provision of public goods hinges on the existence of low-cost exclusion technologies: they enable private providers to charge prices and recover the costs of supply. Examples of jointly consumed goods that have been often provided by private suppliers include roads, bridges and other transport facilities.

When exclusion costs are low, “clubs” and other organisations can also provide public goods for their members (see Buchanan 1965; Olson 1965). Common examples of *club goods* include industry-wide marketing and quality assurance schemes such as marketing labels for organically grown produce. Horizontal price cooperation between firms in the form of cartels is another albeit illegal example. A neighbourhood watch is an example of self-provisioning of public safety.

### **Definitions and Analysis of Externalities**

Marshall discussed “external economies” in his *Principles of Economics* in the end of the 19th century, so externalities appeared in the economic discourse around the same time as public goods. Marshall considered that the expansion of markets, technological development and improved economic infrastructure conferred benefits to agents – benefits which were created by decisions made by others. He defined external economies as those which are dependent on the general development of industry, in contrast to internal economies which depended on the firm’s own resources, organisation, and management practices (Marshall 1947:266). In essence, external economies lowered marginal costs over time.

The contemporary understanding of externalities can be traced back to Pigou’s seminal work on welfare economics. He (1929:176) argued that divergences between

marginal private and social net products reduce the national dividend from its potential. These divergences could occur between agents and their tenants, agents and the general public, and agents and those who purchase goods or services from them. The first and third instances of Pigou’s external economies are today usually referred to as principal-agent problems and imperfect information. The present conception of externalities relates to Pigou’s second type of divergences, which he defined as instances where “person A, in the course of rendering some service, for which payment is made, to a second person B, incidentally also renders services or disservices to other persons C, D, and E, of such sort that technical considerations prevent payment being exacted from the benefited parties or compensation being enforced on behalf of the injured parties” (1929:185).

Pigou has been branded the father of market failure paradigm. He indeed argued that “in any industry, where there is reason to believe that the free play of self-interest will cause an amount of resources to be invested different from the amount that is required in the best interests of the national dividend, there is a *prima facie* case for public intervention” (1929:331). He suggested that taxes could be used when external diseconomies are present while external economies would warrant the use of subsidies. Yet he reminded that divergences are only a *prima facie* case for intervention and that the government cannot necessarily improve the situation because it is plagued by perverse incentives, incompetence and rent-seeking (1929:332-335).

Despite some early reactions to Pigou’s work (see Papandreou 1994), Baumol could still conclude in the early 1950s that “the fact that categories like ‘external economies’ and ‘diseconomies’ remain largely empty

economic boxes prevents any further application of welfare theory as it now stands” (1952:207). However, externalities received more attention during the 1950s (Baumol 1952; Scitovsky 1952; De V. Graaff 1957). The core definition of externalities suggested by Pigou—a side effect of the activity of one agent impacting others and which is not paid or compensated for—was endorsed by a new generation of economists (Kapp 1950; Baumol 1952; Bator 1958; Mishan 1971). However, as was with public goods, references to externalities became more ambiguous. For example, Meade (1952) and Bator (1958) explicitly blurred the borderline between externalities and public goods. Meade (1952) distinguished between two types of externalities he called “unpaid factors” and “atmosphere”, which corresponded with ownership failures and public goods. Bator (1958) in turn defined externalities even more broadly, effectively equating them with market failures. He distinguished between ownership, technical, and public good externalities when trying to trace their origins to institutional failures, indivisibility of goods, and joint consumption. These ambiguities tell of the difficulty to define externalities consistently.

Externalities can be defined consistently, however. The difficulty of excluding users from certain goods or bads makes it impossible to establish effective rights to them and to charge prices or to exact compensation. If goods that are difficult to exclude from are jointly consumed, they are public goods as already discussed. When goods that are difficult to exclude from are subject to *rival consumption*, we have externalities at hand.

However, it is not a trivial task to determine the nature of goods or sometimes even what are the pertinent goods. For

example, it may appear natural to consider air quality as a jointly consumed good. However, it would be more appropriate to consider air basins as the pertinent goods. Air basins have multiple uses and have a finite capacity to cater for users in most uses. When the use of air basins as waste sinks reduces the ability of others to use the air basin, the uses are rival. The governance challenge is that it is difficult to prevent unauthorised use of air basins as waste sinks. Today we face a similar problem with global atmospheric sinks for carbon dioxide.

A number of economists understood that “externalities” are created by the difficulty of exclusion and the resulting lack of effective rights of ownership. Perhaps most influentially, Coase (1960) argued that if one follows the Pigovian reasoning which ignores transaction costs (the costs of using the market system, which include the costs of seeking information, negotiating, making contracts and enforcing them) all that is needed is to establish private property rights and to allocate them to one of the involved parties. The parties would then be able to reach the efficient allocation of resources by bargaining without government intervention. The implication was that resource allocation is independent from distributive choices: different distributions would result in the same efficient outcome (see Samuels and Medema 1998). This is the essence of what is usually called “the Coase Theorem”. However, it could also be renamed “the Stigler Theorem” because it is essentially Stigler’s (1966:110-114) interpretation of Coase’s argument on externalities.

Coase’s own argument was broader than Stigler’s influential interpretation of it. Coase (1960) criticised Pigou’s recommendation of government intervention such as the imposition of fees and the paying of subsidies as a response to externalities.

He showed how Pigou's own analytical approach did not warrant such a recommendation (Samuels and Medema 1998). However, he also demonstrated how the lessons change when transaction costs are recognised, arguing that "there is no reason why ... governmental administrative regulation should not lead to an improvement in economic efficiency" (Coase 1960:18). He demonstrated that the specification and assignment of rights are important policy issues because alternative definitions of rights will result in different economic outcomes in the presence of transaction costs.

However, Coase made the same mistake he accused Pigou of when arguing that allocation of rights should be chosen on the basis of cost-benefit considerations (see Samuels and Medema 1998). He was more correct when suggesting that "the choice among different social arrangements ... must ultimately dissolve into a study of aesthetics and morals" in the presence of transaction costs (see Coase 1960:43; see also Calabresi 1991).

Coase's arguments suggest that externalities prevail because transaction costs prevent the establishment of effective entitlements to all goods (and bads). They also suggest that transaction costs influence what governance solutions will be used, and that these may not be private property rights and markets. Attenuated systems of rights can perform governance functions well in the presence of transaction costs although they may not appear optimal in the analytical frameworks that ignore transaction costs. For example, Hardin (1968) argued that common ownership of natural resources such as forests, pastures and fisheries will lead to their destruction or "the tragedy of the commons". This prediction provided the

rationale for nationalisation and privatisation of natural resources in the developing world.

However, substantial amount of evidence exists now according to which common ownership can facilitate sustainable use of resources where open access would lead to their destruction and private ownership is not a feasible alternative (Bromley and Cernea 1989; Ostrom 1990). Ironically, misinformed nationalisation and privatisation caused wide-spread environmental destruction in the developing world because they destroyed the security of customary entitlements and replaced them with uncertain and illegitimate entitlements which could not be enforced effectively.

Others have taken Coase's observations to the extreme, arguing that externalities exist solely because of transaction costs—or, more specifically, because of exclusion costs—and that externalities may reflect what it is most efficient to do when exclusion costs are properly considered (Buchanan and Stubblebine 1962; Demsetz 1964). Dahlman (1979) has demonstrated that, when taken to its logical conclusions, this line of reasoning leaves no conceptual ground for suggesting government intervention into externality situations. Yet this does not mean that we should not intervene into externality situations: it simply means that the conventional externality scholarship cannot provide consistent justifications for doing so.

Dahlman (1979) himself suggested that we should look carefully at the consequences of institutional alternatives in an externality situation. Different institutional arrangements define entitlements differently and will result in different levels and distributions of transaction costs. Therefore, they will also influence what efficient outcomes will be produced (see also Calabresi 1991; Paavola

2002; Samuels and Medema 1998). This is in line with the above cited view of Coase according to which institutional choices are ultimately a matter of aesthetics and morals.

## Towards the Theory of Governance

Concepts of public goods and externalities have been widely used in policy analysis although they are ambiguous, do not facilitate detailed analysis of policy problems, and do not shed light on the range and implications of institutional arrangements that can be used as governance solutions. These shortcomings have become more problematic in the age of governance, because governance solutions are becoming more complex and have to balance between and realize multiple goals. Yet this does not render the literature on public goods and externalities useless. Quite the contrary, its re-interpretation can form a broader conceptual basis for governance.

The scholarship on public goods and externalities has made several important observations that can support a broader theory of governance. For example, it is understood that externalities and public goods are instances where agents are interdependent (see Baumol 1952:24; Scitovsky 1954:144). *Interdependence* exists when agents' choices or alternatives depend on choices made by others. In the standard economic analysis agents are thought to make their choices independently of choices made by others. However, interdependence is rather an exception while interdependence is pervasive. One person's consumption of most goods influences the choices of others (see Paavola, 2001). This creates conflicts – such as who gets to consume the goods – which have to be resolved in one way or another. Resolution of conflicts is the substance of governance and institutional arrangements are its instruments (Young

1994: 15). Several institutional solutions exist for resolving any particular conflict and they usually have different consequences.

The second important observation of the public goods and externalities scholarship is that goods have attributes that create and shape interdependence. That is, interdependence does not come in one form only: the nature of interdependence varies according to the attributes of involved goods and resources. This observation is the key to detailed analysis of governance problems and to the identification and assessment of governance solutions. Rival and joint consumption and the level of exclusion costs are among the central attributes that shape interdependence. However, there are also other attributes, such as economies of scale, transaction costs and fluctuations in the yields or levels of service that create and shape interdependence (see Schmid 1987). Moreover, the source of interdependence can lie in the attributes of agents as Amartya Sen's argument on the impossibility of a Paretian liberal demonstrates (Sen 1970).

Consumption attributes and the level of exclusion costs divide goods into four main categories (Randall 1983). See Table 1.

*Table 1: Attributes and Types of Goods.*

	Rival Consumption	Joint Consumption
Low Exclusion Costs	<u>Private goods</u> private supply, pricing public supply, exclusive access public supply, equal access	<u>Toll &amp; club goods</u> self-provisioning private supply, pricing private supply, equal access public supply, exclusive access public supply, equal access
High Exclusion Costs	<u>Common-pool goods</u> private monopoly self-governance government regulation	<u>Pure public goods</u> public provision

Goods with rival consumption and low exclusion costs—pizzas, shoes, and arm chairs—are usually referred to as *private goods*. Goods of rival consumption and high exclusion costs are *common-pool resources*. Typical examples of them include pastures, fisheries, aquifers, and oil fields but they also include air basins, watercourses, and atmospheric sinks. Joint consumption and low-cost exclusion characterize *toll goods* and *club goods* which include roads, bridges and local amenities, for example. Finally, jointly consumed public goods which have high exclusion costs, such as public safety and defence, are called *pure public goods*. However, goods may not fit neatly into these categories. Some resources such as bodies of water have multiple uses, some of which are joint while others are rival.

Each good type creates different kind of interdependences and poses different kind of governance challenges. The governance problem with private goods is: whose claims to them are to be recognized and how. In market economies this governance problem is often resolved by the establishment of private property rights and markets, which create incentives that are conducive to the satisfaction of preferences and economic growth. Yet they are by no means an exclusive solution for the governance of private goods. In many market economies private goods such as education, health care and child care services are publicly provided. One reason for this is that public provision can achieve more equal or fair distribution of goods than markets, which distribute goods according to ability and willingness to pay. Direct intervention into the distribution of merit goods is sometimes considered preferable to cash transfers, because people would not consume enough merit goods even if their ability to pay was improved. Yet while the access to merit

goods may be broader under public provision than it would be under markets, it is still often exclusive. Certain level of exclusivity can also often be justified by the scarcity of resources that are needed for the provision of goods and by its effect on individual incentives.

The use of common-pool resources (CPRs) is rival just like that of private goods, but it is more difficult to exclude unauthorised users from CPRs. Rival use again creates the interdependence which calls for the establishment of entitlements. However, it would be difficult to enforce private property rights because the difficulty of exclusion creates opportunities and incentives for unauthorized use. For this reason, private property rights are not often established in the first place.

There are several ways to deal with the interdependence created by common-pool resources, however. One is to establish private monopoly: the incentive for free riding disappears when only one user remains. The second alternative is for the state to establish regulations that specify the rights and duties of involved users with regard to the resource. This has been the governance solution for environmental media, for example. The third alternative is for the involved parties to devise a system of self-governance that specifies the responsibilities and entitlements with regard to the resource. Natural resources such as forests and fisheries are governed by communities and organised user groups in this fashion both in the developing and developed world (Ostrom 1990). Nation states have adopted comparable solutions—international environmental conventions—for the governance of transboundary and global environmental resources (Young 1994). All these institutional solutions replace market

transactions with administrative ones. In monopoly, transactions are internal to the firm while in the case of self-provisioning they are internal to the involved organisation. In the case of regulation, administrative transactions take place within the state.

Joint consumption makes users interdependent because goods such as biodiversity, landscapes and heritage are available for all when they are available to one agent. The agents thus have an incentive to ride free on the effort of others to provide the good. If nothing is done to constrain free-riding, nobody will make effort to provide the good. Moreover, the same quantity and quality of good is available to all. Thus the conflict is: how much of the resource and of what quality should be provided when preferences differ and how should the costs of provision be distributed? Joint consumption creates an incentive to misrepresent one's preferences in these choices in anticipation that one can use the good for free when it becomes available. Governance institutions have to resolve these issues in one way or another with attendant distribution of burdens.

When the costs of excluding users from jointly consumed goods are low, cost recovery is technically easy. Yet the methods of cost recovery distribute costs differently and cause conflicts among the involved agents. Recall that there is no single price that would result in the optimal level of joint consumption. Perfect price discrimination—charging each user a different price, one which equals their personal willingness to pay for the use of good (including letting those not willing to pay to use the good for free)—would result in optimal consumption. However, perfectly discriminating systems are unattainable in the world of positive transaction costs,

because users have incentives not to disclose their willingness to pay. Some cost recovery scheme must be used and it will enable some users to use the good at less than their willingness to pay. When involuntary payments such as taxes are used for cost recovery, some will in turn pay more than they would be willing (see Brubaker 1975; Schmid 1987).

The interaction of agents in the context of involved good attributes and chosen institutional arrangements will generate a particular set of outcomes. Governance outcomes can be predicted by conducting a comparative institutional analysis in the light of involved good attributes. While this approach generates observations that are useful for the design and choice of governance solutions, it is not necessarily easy. Pinpointing the relevant set of attributes that shape the governance problem is far from trivial, as the difficulty of past scholarship to get to the bottom of externalities and public goods demonstrates. Institutional arrangements in turn come in many forms and designs. Perhaps more than anything else, this suggests that policy analysis should be an *Art* with a broad focus and sensitivity to particularities and context. This contrasts strongly with the past view of policy analysis as a *Science*, which in its reductionism could only give a false feeling of exactness.

## Conclusions

The scholarship on public goods and externalities has fundamentally shaped public finance and policy in the 20th century. It has provided the rationale for the welfare and regulatory state and has guided public policies from primary education to nature conservation and military armament. However, its lack of precision requires reinterpretations that seek to guarantee the

usefulness of notions of public goods and externalities during the era of governance.

That said, scholarship on public goods and externalities can contribute to a more general approach to governance. It offers several concepts that give a firm grip on the attributes of goods, the interdependences they create, and the necessity of adopting institutional arrangements to govern those interdependences. The research on public goods and externalities also gives tools to analyse the consequences of adopting particular institutional arrangements as governance solutions.

### Selected References

- Bator, Francis. (1958) "The Anatomy of Market Failure", *Quarterly Journal of Economics*, Volume 72, pp. 351-379.
- Baumol, William J. (1952) *Welfare Economics and the Theory of the State*. Cambridge: Harvard University Press.
- Bromley, Daniel W., and Michael M. Cernea. (1989) *The Management of Common Property Natural Resources: Some Conceptual and Operational Fallacies*. Washington DC: World Bank, Discussion Paper 57.
- Brubaker, Earl R. (1975) "Free Ride, Free Revelation, or Golden Rule," *Journal of Law and Economics*, Volume 18, pp. 147-61.
- Buchanan, James M. (1965) "An Economic Theory of Clubs", *Economica*, Volume 32, pp. 1-14.
- Buchanan, James M. (1968) *The Demand and Supply of Public Goods*, Rand McNally & Co, Chicago.
- Buchanan, James M. (1972) "Who Should Pay for Common-Access Facilities?", *Public Finance*, Volume 27, pp. 1-8.
- Buchanan, James M. and William Craig Stubblebine. (1962) "Externality", *Economica*, Volume 29, pp. 371-384
- Calabresi, Guido. (1991) "The Pointlessness of Pareto: Carrying Coase Further", *Yale Law Journal*, Volume 100, pp. 1211-37.
- Cheung, Steven. (1970) "The Structure of Contract and the Theory of a Non-Exclusive Resource", *Journal of Law and Economics*, Volume 13, pp. 49-70.
- Cheung, Steven. (1973) "The Fable of the Bees: An Economic Investigation", *Journal of Law and Economics*, Volume 16, pp. 11-33.
- Coase, Ronald H. (1960) "The Problem of Social Cost", *Journal of Law and Economics*, Volume 3, pp. 1-44.
- Coase, Ronald H. (1974) "The Lighthouse in Economics", *Journal of Law and Economics*, Volume 17, pp. 357-376.
- Cowen, Tyler. (1988) "Public Goods and Externalities: Old and New Perspectives", in Tyler Cowen (Editor), *The Theory of Market Failure: A Critical Examination*. George Mason University Press, Fairfax, pp. 1-26.
- Dahlman, Carl J. (1979) "The Problem of Externality", *Journal of Law and Economics*, Volume 22, pp. 141-162.
- Demsetz, Harold. (1964) "The Exchange and Enforcement of Property Rights," *Journal of Law and Economics*, Volume 7, pp. 11-26.
- Demsetz, Harold. (1970) "The Private Production of Public Goods", *Journal of Law and Economics*, Volume 13, pp. 293-406.
- Ellis, H. and Fellner, W. (1943) "External Economies and Diseconomies", *American Economic Review*, Volume 33, pp. 493-511.
- Field, Barry C. (1989) "The Evolution of Property Rights", *Kyklos*, Volume 42, pp. 319-345.
- Goldin, Kenneth D. (1977) "Equal Access vs. Selective Access: A Critique of Public



- Goods Theory”, *Public Choice*, Volume 29, pp. 53-71.
- De V. Graaff, J. (1957) *Theoretical Welfare Economics*. Cambridge: Cambridge University Press.
- Hardin, Garrett. (1968) “The Tragedy of the Commons”, *Science*, Volume 162, pp. 1241-1248.
- Head, John G. (1962) “Public Goods and Public Policy”, *Public Finance*, Volume 17, pp. 197-219.
- Head, John G. (1974) *Public Goods and Public Welfare*. Durham: Duke University Press.
- Head, John G. (1977) “Public Goods: The Polar Case Reconsidered”, *Economic Record*, Volume 53, pp. 227-238.
- Kapp, K. William. (1950) *The Social Costs of Private Enterprise*. Cambridge: Cambridge University Press.
- Knight, Frank H. (1924) “Some Fallacies in the Interpretation of Social Cost”, *Quarterly Journal of Economics*, Volume 37, pp. 582-606.
- Margolis, Julius. (1955) “A Comment on the Pure Theory of Public Expenditure”, *Review of Economics and Statistics*, Volume 37: 347-349.
- Marshall, Alfred. (1947). *Principles of Economics*. Eighth Edition. London: Macmillan.
- Meade, James E. (1952) “External Economies and Diseconomies in a Competitive Situation”, *Economic Journal*, Volume 62, pp. 54-67.
- Mishan, E.J. (1969) “The Relationship between Joint Products, Collective Goods, and External Effects”, *Journal of Political Economy*, Volume 72, pp. 329-348.
- Mishan, E. J. (1971) “The Postwar Literature on Externalities: An Interpretative Article”, *Journal of Economic Literature*, Volume 9, pp. 1-28.
- Musgrave, Richard A. (1959) *The Theory of Public Finance: A Study in Public Economy*. New York and London: McGraw-Hill.
- Olson, Mancur. (1965) *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press, Cambridge.
- Ostrom, Elinor. (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Paavola, Jouni. (2002) “Rethinking the Choice and Performance of Environmental Policies”, in Daniel W. Bromley and Jouni Paavola (Editors), *Economics, Ethics, and Environmental Policy: Contested Choices*. Malden: Blackwell, pp. 87-102.
- Paavola, Jouni. (2001) “Towards Sustainable Consumption? Economics and Ethical Concerns for the Environment in Consumer Choices”, *Review of Social Economy*, Volume 59, pp. 227-248.
- Papandreou, Andreas A. (1994) *Externality and Institutions*. Oxford: Oxford University Press.
- Pigou, Arthur C. (1929) *The Economics of Welfare*. Third Edition. London: Macmillan.
- Randall, Alan. (1983) “The Problem of Market Failure”, *Natural Resources Journal*, Volume 23, pp. 131-48.
- Samuels, Warren J. and Steven G. Medema. (1998) “Ronald Coase on Economic Policy Analysis: Framework and Implications”, in Steven G. Medema (Editor), *Coasean Economics: Law and Economics and the New Institutional Economics*. Boston, Dordrecht and London: Kluwer Academic Publishers, pp. 161-183.
- Samuelson, Paul A. (1954) “Pure Theory of Public Expenditure”, *Review of Economics and Statistics*, Volume 36, pp. 387-390.

Samuelson, Paul A. (1955) "Diagrammatic Exposition of a Theory of Public Expenditure", *Review of Economics and Statistics*, Volume 37, pp. 350-356.

Samuelson, Paul A. (1958) "Aspects of Public Expenditure Theories", *Review of Economics and Statistics*, Volume 40, pp. 332-338.

Schmid, A. Allan. (1987) *Property, Power, and Public Choice: An Inquiry into Law and Economics*. Second Edition. New York: Praeger.

Scitovsky, Tibor. (1952) *Welfare and Competition: The Economics of a Fully Employed Economy*. London: Unwin University Books.

Scitovsky, Tibor. (1954) "Two Concepts of External Economies", *Journal of Political Economy*, Volume 62, pp. 143-151.

Sen, A.K. (1970) "The Impossibility of A Paretian Liberal", *Journal of Political Economy*, Volume 78, pp. 152-157.

Stigler, George J. (1966) *The Theory of Price*. Third Edition. New York: Macmillan.

Tiebout, Charles M. (1956) "A Pure Theory of Local Expenditures", *Journal of Political Economy*, Volume 64, pp. 416-24.

Turvey, Ralph. (1963) "On Divergences between Social Cost and Private Cost", *Economica*, Volume 30, pp. 309-313.

Vatn, Arild and Daniel W. Bromley. (1997) "Externalities—A Market Model Failure", *Environmental and Resource Economics*, Volume 9, pp. 135-151.

Young, Oran B. (1994) *International Governance: Protecting the Environment in Stateless Society*. Ithaca: Cornell University Press.

Zerbe, Richard O., Jr.; and Howard E. McCurdy. (1999) "The Failure of Market Failure", *Journal of Policy Analysis and Management*, Volume 18, pp. 558-78.

Jouni Paavola  
ESRC Centre for Climate Change  
Economics and Policy  
University of Leeds, Leeds, UK  
j.paavola@leeds.ac.uk

## Reserve Bank Governor or Chair

*Caner Bakir*

### **Introduction**

The governor of a central bank (or chair of the Reserve Bank) is responsible for the administration of monetary policy. The governor's main responsibility is to keep the rate of consumer-price inflation low and stable to accomplish and maintain economic and financial stability. Governors are technically trained experts expected to make decisions informed by virtue of their specialist knowledge and experience.

The majority of central banks are state-owned. Thus, the governor is essentially a government official. Politicians and governors may have different and conflicting policy preferences. On the one hand, elected politicians may seek to win successive elections through expansionary monetary policies (Nordhaus 1975). Politically motivated discretionary monetary expansion can, in turn, generate short-term gains in output and stimulate an economic expansion to reduce unemployment at the expense of price rises (i.e., political business cycle). On the other hand, unelected central bankers do not have to respond to voters' preferences. It is expected that the governors do evaluate monetary policies in a more technocratic manner than do politicians. Accordingly, they are thought to be more conservative than politicians (Rogoff 1985). It is assumed that governors are more averse to inflation (i.e., place a greater weight on price stability) than elected members of the government (Woolley 1984; Goodman 1992).

In times of conflicting views between a governor and politicians, the central bank independence (CBI) *vis a vis* political authority becomes significant (Goodman

1991:329). Central bankers with legal autonomy from the executive and legislative branches of the government are assumed to have the ability to follow objectives that may conflict with these branches. The Central Bank Act and other relevant laws primarily determine institutional forms and structures in which a central banker operates. The CBI may provide an institutional solution to the electoral cycles in the money supply (Kydlund and Prescott 1977; Barro and Gordon 1983). A number of empirical studies suggest a negative relationship between the CBI and inflation rate (Berger et al. 2001). In particular, it has been argued that "legal independence is a significant determinant of inflation in developed countries" (Cukierman 1992:5). Since a government appoints governors and the majority of board members, it can exert influence directly, albeit temporarily, on the central bank's behaviour. Governmental influences may also affect central bank board decisions through government officials or representatives sitting on the board with a vote and/or a veto.

In this way, the proportion of non-governmental appointments and nomination of board members and length of official tenure compared to the electoral cycle are regarded as significant factors affecting the board's decisions. From this perspective, the formal independence of a central banker, which relates to the appointment (e.g., whether the governor and the board members are appointed by the executive branch or not), term (e.g., whether the term of office of central bank governor and the board members exceed the election cycle), and dismissal procedures of a governor and board members legally stipulated in central bank acts, is closely examined to quantify the political clout over central bankers (see

Cukierman 1992; Grilli et al. 1991; Elgie 1998).

### **Chair and Governors of the Fed**

The Federal Reserve System (hereafter the Fed) is the US central bank. The Board of Governors of the Federal Reserve System, Federal Reserve Banks, and Federal Open Market Committee (FOMC) constitute the three main parts of the Fed. The chairman is a member of the Board of Governors (BOG) and the head of the Fed. The chairman and vice-chairman of the Fed are appointed by the president and approved by the US Senate from among the members of the BOG. The chair and vice-chairman have a 14-year term but these positions are based on a separate four-year appointment by the president. A chair who is not reappointed continues to serve as a governor. Thus, in practice, the chair serves only a four-year term.

Evidence from the appointments of the former US president, Lyndon Johnson, shows that the Fed candidates, in Woolley's words, "have been judged by their experience, technical expertise, and acceptability to particular constituencies. ... [i.e.,] bankers, businessmen, or economists" (Woolley 1984:51). For example, in regard to occupational backgrounds of successful candidates in the period between 1955-82, 39 per cent of the all members of the BOG have had backgrounds primarily in the private for-profit sector whereas 36 per cent of all board members have held Federal Reserve (21 %) and financially related posts (15 %) at all levels of government (Woolley 1984:56). Further, 45 per cent of governors worked previously as economists while 60 per cent of board members had some official affiliation with the Fed prior to their appointment (p.57). Almost one-fifth of board members spent the longest period of their career in the Fed itself.

Apart from having experience and knowledge in finance and/or economic related areas, the Fed governors have also common educational backgrounds. For example, 43 per cent of the BOG members studied at one of four elite universities in the US—Harvard, Princeton, Yale, and Stanford (Woolley 1984:58). In regard to the demographic and social characteristics of those at the top of the Fed, Woolley's (1984:67) conclusion in the mid-1980s still seems relevant today: "the Federal Reserve continues a long history of administration by white, male, upper-middle-class technicians."

Woolley also shows that "private sector officials and financial positions did exercise general control over the process of selecting Federal Reserve officials" during the Johnson years (Woolley 1984:55). Recent evidence presented by Jones (1991:52) is in line with this observation; "[m]odern Fed chairmen are chosen, in effect, by the financial markets ... the White House leaks to the press a list of potential nominees for the top job at the US central bank and leaves the choice largely to collective judgement of the financial markets." Hetzel (1990:107) summarises attributes that promote an individual's candidacy for chairman: "a distinguished career of public service, recognition in financial markets, recognition by key politicians (especially the chairman of the House and Senate Banking Committees), acceptability to presidential political advisors." Accordingly, wider political and societal constituencies will accept a Fed chair having these attributes.

The US Federal Reserve's sole decision-making body is the FOMC. It decides by majority vote to change the federal funds rate using open market operations and the purchase and sale of government securities in the secondary markets. The FOMC

usually meets eight times a year to decide on the Fed's open market operations. The US president also appoints a majority of the FOMC members. The FOMC has twelve members of which seven governors are politically appointed by the president with the Senate's approval (partisan appointees), while remaining five bank presidents are selected from among the 12 presidents of Federal Reserve district banks (representative appointees).

The BOG sets the discount rate, the rate at which the Fed lends funds to banks. The BOG appointments are important as they affect monetary policy made by the FOMC. Formally, according to Federal Reserve Act (1935), "financial, agricultural, industrial, and commercial interests and geographical divisions of the country" are considered by the president (quoted in Chang 2001:324). Informally, however, candidate's sex, race, Senate preferences and financial sector opinions are also considered (Havrilesky 1995:291, 296-97; Jones 1995:69).

The seven members of the BOG are nominated by the US president, recommended by the Senate Banking Committee, and confirmed by the US Senate with a 14-year term. If a governor is appointed to a 14-year term but leaves office before the expiration of this term, a new governor is appointed to fill out this unexpired term. He then has to be reappointed to another 14-year term, since he is only appointed for the unexpired time. Thus, the average term is not 14 years. In practice, the governors serve much shorter terms (Beck 1990:124). Further, "presidents get to nominate different numbers of Governors; Carter, for example, had appointed six of the sitting governors by 1980; Reagan had only one such appointment by 1984, although he did have

six appointees by the end of 1987" (Beck 1990:124-25).

Presidents select reliable governors whose views align with their own (Havrilesky 1992). According to Havrilesky and Gildea (1992:397), "the control of Federal Reserve Board appointments provides presidential influence over the direction of monetary policy." Thus, the voting behaviour of partisan appointees of the FOMC reflect the preferences of the president who appointed them (Belden 1989; Gildea 1990; Havrilesky and Gildea 1992). For Havrilesky and Gildea (1992), the U.S. presidents generally make appointments both to influence monetary policy and to appeal to interest groups (e.g., the financial policy community). The literature on the appointment of BOG members contains two competing views. On the one hand, some argue that the president considers the Senate's preferences before the formal nomination of candidates and thus argue for joint influence of the president and the Senate (Calvert, McCubbins, and Weingast 1989; Morris 2000; and Snyder and Weingast 2000). On the other hand, Moe (1985,1997) claims that the president chooses whomever he wants and the Senate appoints the president's candidate. Recently, Chang (2001) argued that politicians influence monetary policy by appointing candidates close to their preferences. However, in her empirical analysis of the bargaining process between the President and the Senate, she shows that the question of whether or not the President and the Senate jointly determine appointees, or one or the other dominates the appointment process is inconclusive. Falaschetti (2002, p.488), in his evidence from a panel of FOMC votes, concludes that: "(1) BOG who were nominated and confirmed by the same party (Republican or Democrat) prefer

significantly looser policy than do other FOMC members; (2) Monetary policy is significantly looser when either party controls the oversight mechanism (i.e., the presidency and Senate) than when control is split; (3) oversight acts less forcefully on district bank presidents than [the BOG] governors.”

In regard to the relationships between the Fed Chair and representative appointees, the 1933 and 1935 Banking Acts of New Deal reforms centralised the decision-making process at the Fed expanding the influence of the Fed chair at the expense of the district bank governors (Maisel 1973).

### **Governors, Turnover Rate and Political Vulnerability**

The turnover rate of the central bank governor (Cukierman et al. 1992; Cukierman 1994) and the political vulnerability of the governor to changes in government (Cukierman & Webb 1995) are used to measure the governor’s informal autonomy to pursue independent monetary policies. Cukierman (1994) shows an inverse relationship between the expected length of tenure of the central bank governor and the inflation rate due to the governor’s effective rate of time preference. In particular, the central bank governor turnover rate (the number of changes of the central bank governor divided by the number of years) is used as the chief behavioural measure of central bank ‘actual’ independence in developing countries (Cukierman 1992:383-86). A higher rate of turnover of central bank governor indicates a lower level of the CBI suggesting lesser freedom to formulate monetary policy independent of government control. The reason is twofold. First, incoming government has the opportunity to make a partisan appointment of a new subservient governor. Second, a new central

banker generally is unknown to markets and bureaucracy, and thus may have less influence than his predecessor until he gains credibility, respect, and clout. Research covering the average turnover rate of central bank governors for developed and developing countries between 1950 and 1989 has shown that the turnover rate differs greatly across countries, varying from a minimum of 0.03 (average tenure of 33 years), for Iceland, to a maximum of 0.93 (average tenure of about 13 months) for Argentina (Cukierman 1992: Table 19.5: 384). The average rate of turnover for the developing countries is 0.28 whereas that of industrial countries is 0.13 (with the highest rate of 0.2, for Japan and Spain).

Although this measure hardly discriminates between central banks of developed countries, there is a very strong association between inflation rates and the turnover rates of central bank governors in high inflation developing countries (Cukierman 1992; Cukierman et al. 1992). For example, the average actual term of office of the governor in Argentina was about ten months while average yearly inflation was 319 percent in the 1980s; this in spite of the fact that the legal term of office of the governor is four years. This is partly due to a tradition in Argentina that the governor offers his resignation to a new government (Cukierman 1992:383).

During the 1990s, central bankers in Latin America took significant steps towards legal political independence from governments due to central banking reforms (Jacome 2001). For example, the term of office for governors either overlaps or exceeds the election cycle; in most of these countries, the legislative branch is involved in the appointment and/or dismissal of the governors; also some of these countries removed the President’s dismissal power

over governors. However, in spite of legal reforms, governor turnover rates in these countries are still higher than that of developed countries. (Argentina had its fifth central bank governor in less than two years between 2000 and 2002.)

The probability of the central bank governor changing due to political instability is measured by the degree of political vulnerability of the central bank governor (the number of replacements of the central bank governor within 6 months of a political transition, i.e., shift in regime or change in the head of government divided by the number of political transitions). Cukierman (1994) investigated two types of political instability; (i) party political instability (the frequent change in government between political parties) and (ii) regime political instability (radical change in political regime due to revolution or *coup d'état*). He demonstrated that party political instability has a positive correlation whereas regime political instability has a negative correlation with the CBI.

Cukierman and Webb (1995) compare the probability of a central bank governor being replaced after a democratic and non-democratic change of government to show how turnover of the governor relates to political events. They found that the probability of the governor's replacement after non-democratic change in a government (a coup or revolution) is much higher than a regular government change. In particular, they show that a governor will be replaced about half the time within six months of a non-democratic change whereas the governor is much less likely to be dismissed within six months following a democratic change in the head of government – “about one-fourth of the time in developing countries and one-tenth in industrial countries” (Cukierman and Webb

1995:397). For the period from 1950 to 1989, they calculate that the average of the political vulnerability index is 0.24 (0.10 for industrial countries; 0.34 for developing countries).

### **Fed Governors and Leadership**

Apart from formal and informal autonomy of central bank governors, the leadership quality of the governors may affect the behaviour of a central bank. As such, Friedman states that “... the extent to which a system of this kind is really a system of rule by man and not by law and is extraordinarily depended on the personalities involved” (Friedman 1962:235). From this perspective, Kettl (1986:193) argues that “throughout the Fed's history, its power over the economy has depended more on the political leadership of its chairman than on any other factor [e.g., CBI].” Two former Fed governors, Arthur Burns and Paul Volcker are cases in point. Burns adopted politically motivated expansionary policy in 1972 which was in accord with the policy goals of the then US president Richard Nixon. Conversely, Volcker pursued the tight monetary policy which was detrimental to the re-election of the then US president Jimmy Carter in 1979. For Jones (1991:89-90), “great Fed leaders are distinguished more by their policy instincts, timing, effective and credible communications, and performance under pressure than by their technical monetary competence and number crunching ability, though the latter should not be minimised.” Following this theme, among the contemporary Fed chairmen, Jones (1991:ch.3) considers Alan Greenspan (1987-2004), Paul Volcker (1979-1987) and Marriner Eccles (1934-1948) as the great Fed chairmen. However, although acknowledging that the personality of the governor matters, Siklos (2002:85) argues

that “the primacy of personality of the governor is more apparent at times of crisis than as a feature of the entire term of a particular central bank [governor].”

### **Governors and European Central Bank**

The case for formal autonomy of central bankers is based on the assumption that they are not perfect agents of their political principles because they are conservative (i.e., inflation-averse). However, as Lohmann (1998a:360) puts it, “[i]n the extreme, central bankers who are perfect agents would voluntarily accommodate electoral pressures, no matter how much formal independence they are granted.” Thus, “[p]luralism of appointing bodies may serve to deflect political influences on monetary policy” (Lohmann 1998a:361; see also Lohmann 1998b). For example, the European Central Bank (ECB), mirroring the Deutsche Bundesbank (the German central bank) has decentralised appointment bodies insulating the conduct of monetary policy from political influences (Lohmann 1997).

The ECB acts in conjunction with the European System of Central Banks which is composed of 15 European Union member states’ central banks and a six-member Executive Board. The ECB is run by a Governing Council and an Executive Board. In Europe, all the EU member states must agree on appointments to the Executive Board. The Council of Economics and Finance Ministers, which is composed of the finance ministers of the EU members, recommends the members of the Executive Board. Then, the European Parliament (one of the legislative institutions of the EU) and the Governing Council of the ECB (which consists of the members of the Executive Board and the heads of the euro area member countries’ 12 central banks which exclude Denmark, Sweden and the United

Kingdom) are consulted. Finally, the heads of state of the euro area members confirm the appointments for non-renewable eight-year term (for a comparison between the ECB and the Fed, see Pollard 2003).

### **Governors and Global Governance**

Central bankers can be seen as a group of technocrats with common identity, interests, orientations and ideas. As Paul Volcker, a former chairman of the US central bank (i.e., Federal Reserve) observed, central bankers have common “experience, tenure, and training”, and hence they “are almost uniquely able to deal with each other on a basis of close understanding and frankness” (cited in Helleiner 1994:200). As Johnson observes “by the late 1980s, central bankers (both from individual central banks in Western Europe and North America, and from the international financial institutions) had effectively coalesced into a transnational epistemic community” (Johnson 2002:14).

As such, the Basle Committee on Banking Supervision is now an example of a transgovernmental network among central bankers; 13 central bank governors of developed countries sitting on the board of the Committee are members of what Coleman and Perl (1999) would call a ‘transnational expert community.’ These governors are an integral part of global governance in finance as they have the collective capacity to identify and solve some financial problems on a global scale. Indeed, they play pivotal role in the development of codes of best practices in banking supervision. Further, due to their overlapping memberships they mediate between the Basle Committee and their domestic banking community as ‘policy entrepreneurs’ coupling problems with policy solutions and political processes at both domestic and transnational levels. Also,



their advice on financial regulatory issues is crucial. As Pierce (1990:153) observes in the US context, “the chairman’s position on any type of economic legislation, either domestic or foreign, often is crucial in determining the fate of a specific bill.”

Further, since the 1980s, there has been such convergence of central banking ideas among the central bankers of the developed countries that a broad consensus was formed around the so-called two key principles of central banking: price stability and the CBI (Blinder 1999; Zulu et al 1994:131). Johnson (2002:2) argues that these two key principles as well as “basic techniques and practices, and internal organisational frameworks” are transmitted into domestic policies of post-communist countries through the “active, extensive training and technical efforts of the Western central banking community” to create Western-style central banks in post-communist states. In a similar vein, Marshall (1999) shows the significant influence of a group of central bankers on the policies and on the ultimate design of the ECB.

### **Accountability and Transparency**

In order to assure that central bankers follow conservative monetary policy, they can be committed to a low-inflation monetary target by a central bank act (Rogoff 1985).

This can be coupled with a central banker’s contract which refers to explicit inflation or monetary aggregate targets. The New Zealand Reserve Bank Act of 1989, for example, refers to both ‘inflation targeting’ and ‘contract’. It is recognised that governors are held responsible for a failure to meet monetary policy targets. For example, the Policy Targets Agreement, which forms the basis of the accountability of the governor in New Zealand, requires that the governor of the Reserve Bank of New Zealand explain why inflation has

deviated outside the specified range and what measures it will take to ensure that it moves back inside. The governor may be dismissed if an inflation target is breached. Accountability of governors helps to ensure that the actions of the governors are in line with a commitment to price stability.

The independence of an unelected central bank governor *vis a vis* democratically-elected government may create a ‘democratic deficit.’ Thus, the accountability of central bankers becomes a significant issue in the governance of central banking. From this perspective, a central bank governor’s appearance before legislative bodies (e.g., the semi-annual Humphrey Hawkins testimony by the chair of the Fed) and their report to government on a regular basis have also become key components of a governor’s accountability; this may also be stipulated in legislation, as in the Bank of England Act and the legislation of the ECB (see Paulin 2000:9; Jacome 2001; Bollard 2003).

Central bankers also publish periodical financial statements as well as press releases and conferences to inform the public about the activities of their central banks. In doing so, the transparency of the central bank and its public communication are facilitated. In the US, for example, the Fed chairman has the official authority to speak for the Fed. The chairman also speaks before the US Congress, the Executive branch, and the public.

Although the accountability of Latin American central bankers increased following the central banking reform of the 1990s, “transparency of policy formulation... and... disclosure of central bank financial statements is still weak” (Jacome 2001:12). Argentina, Bolivia, Brazil, Chile, Colombia, Honduras, Mexico, Uruguay, and Venezuela have sound

accountability requirements. For example, ‘the governor of the central bank must inform executive branch and congress, or a specialised committee, of the results of the central bank’s monetary policy and other activities, and the extent to which policy goals have been achieved’ (p.12). However, in other Latin American countries, notably in Nicaragua, “legislation merely requires the publication of an annual report” (p.12). In Costa Rica, Dominican Republic, Honduras, and Guatemala, central bank reforms do not make notable reference to “the transparency of financial statements, and hence, central bank transactions are not always properly recorded” (p.12).

### **Governors on Asset Bubbles**

Until the early 2000s, there was a consensus among central bankers that monetary policy should not be used to curb rapidly rising asset prices such as equities and property. Central bankers take rising asset prices into account only when they boost spending causing future inflation. However, the stock market booms and busts in Japan in the 1980s-2000s and in the US in the 1990s-2000s, both occurred at times of low inflation (or deflation). In particular, an inflated bubble can, in time, cause excessive borrowing and over-investment, which generates financial and economic imbalances. Thus, a growing number of central bankers and economists argue that central bankers may adopt a tight monetary policy in response to economic imbalances caused by climbing asset prices, even if inflation is well within its target range (*The Economist* 7 September 2002, p.72, 18 January 2003, p. 68). Therefore, the case for interest rate rises is stronger when a boom in asset prices is coupled with a rapid growth in credit. Greenspan has been criticised that, in the late 1990s, he should have increased

interest rates in order to curb sharp rises in the prices of assets so as to prevent painful consequences when speculative bubbles burst. Greenspan, however, argued that it was not possible to identify a bubble until after it had burst; also, a modest rise in interest rates would not be effective as it might increase confidence in the central banks’ powers whereas a sharp rise in interest rates could cause a recession (Bernanke 2002). As a result, it has been argued that lowering interest rate to limit the economic damage after a stock market bust is preferable to pop asset bubbles through interest rate rises. Consequently, central bankers may also have different views on whether they should respond rapidly to rising asset prices by pricking the bubble before it burst or give it a chance to deflate on its own.

### **Internet Sites**

Speeches of federal Reserve Officials.  
[www.federalreserve.gov/boarddocs/speeches/2009/default.htm](http://www.federalreserve.gov/boarddocs/speeches/2009/default.htm)  
European Central Bank. [www.ecb.int/](http://www.ecb.int/)  
Bureau of International Settlements.  
[www.bis.org/review/review.htm](http://www.bis.org/review/review.htm)

### **Selected References**

- Barro, R. and D. Gordon. (1983) “Rules, Discretion, and Reputation in a Model of Monetary Policy”, *Journal of Monetary Economics*, Volume 12, pp. 101-21.
- Beck, Nathaniel. (1990) “Political Monetary Cycles”, in T. Mayer (Editor), *The Political Economy of American Monetary Policy*. Cambridge: Cambridge University Press, pp 115-130.
- Belden, S. (1989) “Policy Preferences of FOMC Members as Revealed by Dissenting Votes”, *Journal of Money Credit and Banking*, Volume 21, pp. 432-41.
- Berger H; J. de Haan; and S. Eijffinger. (2001) “Central Bank Independence: An Update of

- Theory and Evidence”, *Journal of Economic Surveys*, Volume 15, pp. 3-40.
- Berger, H. and U. Woitek. (1997) “How Opportunistic Are Partisan German Central Bankers?”, *European Journal of Political Economy*, Volume 46, pp.17-39.
- Bernanke, B.S. (2002) “Asset-Price “Bubbles” and Monetary Policy”, in *The New York Chapter of the National Association for Business Economics*. New York: NABE.
- Blinder, A.S. (1999) *Central Banking in Theory and Practice*. Cambridge, Mass.: MIT Press.
- Bollard, A. (2003) “Corporate Governance in the Financial Sector”, *The Annual Meeting of the Institute of Directors*. Christchurch, New Zealand. 7 April.
- Calvert, Randall L; Matthew D. McCubbins; and Barry R. Weingast. (1989) “A Theory of Political Control and Agency Discretion”, *American Journal of Political Science*, Volume 33, pp. 588-611.
- Chang K. (2001) “The President Versus the Senate: Appointments in the American System of Separated Powers and the Federal Reserve”, *The Journal of Law, Economics, and Organization*, Volume 17, pp. 319-55.
- Cukierman, A. (1992) *Central Bank Strategy, Credibility, and Independence: Theory and Evidence*. Cambridge, Mass: MIT Press.
- Cukierman, A. (1994) “Central Bank Independence and Monetary Control”, *Economic Journal*, Volume 104, pp. 1437-1448
- Cukierman, A; S. Webb; B. Neyapti. (1992) “Measuring the Independence of Central Banks and Its Effect on Policy Outcomes”, *World Bank Economic Review*, Volume 6, pp. 353-98.
- Cukierman, A. and S. Webb. (1995) “Political Influence on the Central Bank-International Evidence”, *World Bank Economic Review*, Volume 9, pp. 397-423.
- Coleman, W. D. and A. Perl. (1999) “Internationalised Policy Environments and Policy Network Analysis”, *Political Studies*, Volume 47, pp. 691-709.
- Elgie, R. (1998) “Democratic Accountability and Central Bank Independence: Historical and Contemporary, National and European Perspectives”, *West European Politics*, Volume 21, pp. 53-76.
- Friedman, B.M. (1962) “Should There be an Independent Monetary Authority?”, in LB Yeager (Editor), *In Search of a Monetary Constitution*. Cambridge, MA: Cambridge, MA.
- Gildea, J. (1990) “Explaining FOMC Members' Votes”, in T. Mayer (Editor), *The Political Economy of American Monetary Policy*. Cambridge: Cambridge University Press, pp 211-28
- Goodman, J. (1991) “The Politics of Central Bank Independence”, *Comparative Politics*, Volume 23, pp. 329-49.
- Goodman, J. (1992), *Monetary Sovereignty*. Ithaca: Cornell University Press.
- Grilli, V; D. Masciandaro; and G. Tabellini. (1991) “Political and Monetary Institutions and Public Finance Policies in the Industrialized Democracies”, *Economic Policy*, Volume 10, pp. 342-92.
- Havrilesky, T. and G. John (1992) “Reliable and Unreliable Partisan Appointees to the Board of Governors” *Public Choice*, Volume 73, pp. 249-61.
- Helleiner, E. (1994) *States and the Reemergence of Global Finance: From Bretton Woods to the 1990s*. Ithaca: Cornell University.
- Hetzel, Robert L. (1990) “The Political Economy of Monetary Policy”, in T. Mayer (Editor), *The Political Economy of American Monetary Policy*. Cambridge: Cambridge University Press, pp 99-114.
- Jacome, H. Luis. (2001) *Legal Central Bank Independence and Inflation in Latin America During the 1990s*. Washington: IMF. Working Papers No 01/212.
- Johnson, J. (2002) *Agents of Transformation: The Role of the West in Post-Communist*

- Central Bank Developments*. Glasgow: Glasgow, pp 1-36.
- Jones, M. David. (1991) *The Politics of Money: The Fed Under Alan Greenspan*. New York: New York Institute of Finance.
- Kydland, F. and E. Prescott. (1977) "Rules Rather Than Discretion: The Inconsistency of Optimal Plans", *Journal of Political Economy*, Volume 83, pp. 473-91.
- Lohmann, S. (1997) "Partisan Control of the Money Supply and Decentralized Appointment Powers", *European Journal of Political Economy*, Volume 13, pp. 225-46.
- Lohmann, S. (1998a) "Federalism and Central Bank Independence: The Politics of German Monetary Policy, 1957-1992", *World Politics*, Volume 50, pp. 401-46.
- Lohmann, S. (1998b) "Institutional Checks and Balances and the Political Control of the Money Supply", *Oxford Economic Papers*, Volume 50, pp.360-77.
- Maisel, Sherman. (1973) *Managing the Dollar*. New York: Norton.
- Marshall, M. (1999) *The Bank*. London: Random House Business Books.
- Moe, Terry. (1985) "Control and Feedback in Economic Regulation: The Case of the NLRB", *American Political Science Review*, Volume 79, pp. 1122-43.
- Moe, Terry. (1987) "Interests, Institutions and Positive Theory: The Politics of NLRB", *Studies in American Policy Development*, Volume 2, pp. 236-99.
- Morris, I. (2000) *Congress, the President, and the Federal Reserve*. Ann Arbor: University of Michigan Press.
- Nordhaus, W. (1975) "The Political Business Cycle", *Review of Economic Studies*, Volume 42, pp. 169-90.
- Paulin, G. (2000) "The Changing Face of Central Banking in the 1990s", *Bank of Canada Review*, pp. 3-12.
- Pollard, S. P. (2003) "A Look Inside Two Central Banks: The European Central Bank and the Federal Reserve", *The Federal Reserve Bank of St. Louis Review*, pp. 11-30.
- Rogoff, K. (1985) "The Optimal Degree of Commitment to and Intermediate Monetary Target", *Quarterly Journal of Economics*, Volume 100, pp. 1169-89.
- Siklos, P. L. (2002) *The Changing Face of Central Banking*. Cambridge: Cambridge University Press.
- Synder, Susan and Barry Weingast. (2000) "The American System of Shared Powers: The President, Congress, and the NLRB", *Journal of Law, Economics, and Organization*, Volume 16, pp. 269-305.
- Vaubel, R. (1997) "The Bureaucratic and Partisan Behaviour of Independent Central Banks: German and International Evidence", *European Journal of Political Economy*, Volume 13, pp. 201-24.
- Woolley, J. (1984) *Monetary Politics: The Federal Reserve and the Politics of Monetary Policy*. New York: Cambridge University Press.
- Zulu, J. B, and I. McCarthy. (1994) *Central Banking Technical Assistance to Countries in Transition*. Washington DC: International Monetary Fund.

Caner Bakir  
 Department of Economics  
 Koç University  
 Istanbul, Turkey  
 cbakir@ku.edu.tr

## Sickness, Absenteeism and Insurance

Daniela Andren

### Introduction

Employee *absenteeism* has been an important subject of psychological and sociological research for a long time, and has received increasingly more attention in the fields of economics and social medicine during the past twenty years especially. *Labor absence* can be thought of as any time spent away from the workplace that is not anticipated or scheduled by the employer. The causes of work absenteeism are a topic of discussion debated from the firm to the macro level. People may be absent from their job because of either their own or another family member's *sickness*, because of death in the family, or for other strictly personal reasons. All these situations imply loss of income for the *individual*: if you do not work you do not get paid. Given that society recognizes that people need to be away from work when sick or injured, a universal solution is to create social insurance to cover individual income loss, for example sickness insurance. However, it may also be the case that there is no universal public commitment, but that individual employers offer various schemes instead.

*Insurance* is a payment with which individuals and businesses protect for a time period themselves, their families, and their property from *possible* losses resulting from unpredictable events such as fires, accidents, and illnesses. *Sickness insurance* is financed by a compulsory payment by employers and/or employees and covers income loss when working capacity is significantly reduced due to sickness. Thus, when individuals are temporary absent from work due to sickness, if there is sickness insurance

individuals get some amount of compensation. This can be in the form of *sick-pay* from the employer or a *sickness benefit* from the public authority. This compensation, which in fact is an insurance for loss of income, is referred to as *sickness insurance* in the rest of this article.

The institutional settings of the sickness insurance may vary greatly: social insurance can take the form of a mandated private insurance or can be provided by a public institution, it can be more or less generous, etc. If there is an insurance that covers labor absenteeism due to sickness, there might be other factors that—given that one is feeling ill—determine absence from work too, such as job involvement and satisfaction, a culture with strict attendance norms, etc. Persons with a high level of job satisfaction, or whose work-culture includes strict attendance norms, may seldom be away due to poor health, whereas low satisfaction, and/or lax norms, may lead to greater absenteeism.

In many countries, public sickness insurance is compulsory. However, often employees are compensated through voluntary agreements, for example between the employer and the employees, or employers' associations and unions. These voluntary agreements may stand on their own or be complements to the public sickness insurance.

Though the purpose of sickness insurance is to pay compensation to individuals whose ability to work is reduced due to illness, both ability to work and poor health are difficult to measure with an objective scale. Therefore, it cannot be ruled out that individuals are absent and receive compensation even though they are not very ill or perhaps not ill at all. In a sense individuals themselves determine the probability of absence. Moreover, general

insurance theory (e.g. Arrow 1974 and Laffont 1989) states that an insurance where the probability of loss can be influenced by the insured should not pay full compensation. Consequently, sickness insurance—and especially publicly administrated sickness insurance—usually pays less than full compensation. Nevertheless, many voluntary agreements (in for example Germany, Great Britain, Finland, and Switzerland) pay full compensation for at least some days.

### **Sickness Absenteeism**

Douglas (1919) was perhaps the first to analyze absenteeism, which was mentioned as “*another factor in the instability of labor*”. Given the fact that during that time it was often confused with labor turnover, Douglas defined absenteeism as “*absence from work at the job at which one is employed*”, including absence for both all-day or only partially. However, an internationally accepted definition of *absenteeism* is *the manifestation of a decision by an employee not to present themselves at their place of work at a time when it is planned by management that they should attend*. Expressed more briefly, it is the non-attendance at work of workers expected to be present. The causes of absenteeism are many and include serious accidents and illness, low morale, poor working conditions, lack of job satisfaction, inadequate leadership and poor supervision, personal problems (e.g., financial, marital, substance abuse, child care etc.), poor physical fitness, transportation problems, a *generous* insurance, etc. Reports on sickness absenteeism at the firm or company level usually define sickness absence as being *the total number of days of sickness absence as a percentage of possible working days*. However, there are several measures of

sickness absence. For example, Hensing et al. (1998) reviewed previous studies of sick leave and suggested five measures when analyzing sick-leave data within an *epidemiological* framework: frequency of sickleave, length of absence, incidence rate, cumulative incidence and the duration of a sick-leave spell.

Although the burden is not equally divided among the different parts that contribute to the coverage of labor absenteeism (e.g., employees, employers, insurance companies, and governments), they all bear a part of the *costs* related to absenteeism. Without compensation, the *employee* would have reduced income as result of absenteeism related to ill health. In addition, he or she could have additional expenditures related to health care services, but also a loss of welfare in the form of pain and/or suffering. For the employer, absenteeism implies both direct and indirect costs, or, more exactly, *financial costs* (e.g., sick pay, lost productivity, inferior quality, and the wage costs of replacement employees) and *administrative costs* (e.g., labor costs to hire replacement employees or to re-assign the remaining employees, and labor cost to maintain and control absenteeism). *Insurance companies* may insure both the absenteeism risk and the health of the workers and their families. Usually they have to pay the benefits in case of absenteeism and the health care costs of the ill health of the employees.

The *national economy* is affected through a loss of potential output from the reduction in labor supply. This also affects the *national budget*, through tax revenue losses, and increased expenditures, if the payments of sickness benefits and costs of health care and medical treatment are covered by a public system.

The definition of absenteeism, its causes, and its costs in terms of finances and administrative effectiveness are quite clear, while ways to take affirmative action to control absenteeism, costly administration and misuse are far from clear-cut. Traditional methods of absenteeism control based only on disciplinary procedures have proven to be ineffective. *Discipline* alone usually does not identify or address the root causes of absenteeism, and can, at best, give the illusion of control. However, through skilled managers and supervisors, management can aim to develop a willingness on the part of employees to attend work regularly and to assist in motivating coworkers.

### **Sickness Insurance**

Many industrialized countries have *social insurance* schemes that encompass sickness benefits. Sickness benefits are cash transfers related to *loss of earning because of the temporary inability to work due to illness*. This excludes paid leave related to sickness or injury of a dependent child that is recorded under family cash benefits. *All expenditure on public and private health care are recorded under health*.

A common misconception about *income protection* plans is that they are a benefit as are vacations, and as such, should be fully utilized. The truth is income protection plans are *an insurance*. The sole and only purpose of pay for sick leave is to assist in protecting employees against loss of income in the event of an unavoidable absence due to sickness or a non-work related injury. Use of income protection plans for any other purpose negates their intent.

### **Salient Facts**

The first general social insurance scheme was introduced in the last two decades of the

1800s: in Germany, Chancellor Otto von Bismarck's 1883 sickness insurance law provided to employees in defined types of industry both medical care and cash benefits during a period of sickness, to be paid for out of contributions from both employees and employers. Austria followed part of the German example in 1888, Italy in 1893, and both Sweden and The Netherlands in 1901. According Bismarck's law, *a proportion of previous earnings were to be paid in cases of sickness, injury, widowhood, and old age*.

Sickness absence is covered by social security in many countries (especially Europe), and the long-term consequence of sickness—disability—is covered in just about all industrial countries. There was relatively widespread complacency about rising expenditures on especially disability among the major players in most European countries until the end of the 1980s and in the 1990s. If social security is financed with contributions levied on wages, then there is an effect on labor costs and on the ultimate price of products. This implies pressure on the governments of the countries with a generous social insurance system.

Given that the institutional settings vary across countries, one way of comparing the governments' payments for sickness benefits is as the percentage of these payments in GDP, or in the public social expenditure.

The richest sources of data on sickness absenteeism across countries over the years are the Labour Force Surveys and Household Surveys. Barmby et al. (2002) show how internationally and intertemporally consistent information on sickness absence can be constructed from Labour Force Surveys, and describe some important features of data that they have generated using the Luxembourg Employment Study.



A good comparison across countries requires a good definition of sickness absenteeism. Since Labor Force Surveys use the same technique regardless of country, and are not affected much by the institutional settings, e.g., whether insurance is partially or wholly financed by the public sector, they provide the best available source for comparison.

OECD Health Data report *absence from work due to illness* since 1960 for several countries, but there are missing information for several countries for some years. Absence from work due to illness is defined as *the number of workdays lost per year per currently employed person*.

### Previous Studies

The problem of employee absenteeism has long been an important subject of psychological and economic research and modeling. In both, individuals are often assumed to decide daily to attend the work or non-work alternative with the highest expected utility. The conventional labor supply model of absence focuses on the role of contractual arrangement, assuming that the wage rate plays a central role in the decision to work or not to work. There are other economic factors that might influence this decision, such as the replacement rate, the tax rate, and employee sharing plans (e.g., profit-sharing and/or employee share-ownership).

Alexanderson (1998) examines carefully about 320 studies of sickness absence, selected by searching some bibliographic reference databases, for the period 1988-1997. Unfortunately, due to the selection of these databases, the analysis does not include some seminal economic studies. Even with this limitation, we learn several good lessons from this examination. Firstly, the analyzed studies distinguished mainly

between two types of absenteeism. These two are often described in terms of various dichotomies, such as avoidable/unavoidable; voluntary/ involuntary; unrelated to ill health/related to ill health; uncertified/certified; unauthorized/authorized; non-granted/granted; unnecessary/necessary; etc. The term *sickness absence* is used to refer to the latter half of each dichotomy; in other words, it is related to *inability rather than unwillingness to work*, but many of these studies is concerned with the first part. Secondly, the relationship between different factors and sickness absence is analyzed at three levels: national, workplace or communal, and individual level. And thirdly, the reviewed studies lack a medical perspective.

Most of the previous literature in economics analyzes the relationship between health, health insurance, and labor outcomes (see Currie and Madrian 1999 for a survey), and not much is known about sickness absenteeism, sickness insurance and labor outcomes. From the economic literature about sickness absenteeism, we learned that despite these differences in average levels of sickness absenteeism across countries, patterns of absence according to age and sex are very similar. Male absenteeism is generally lower than female. Single men have the lowest absence rates while married women have the highest absence rates. The older the age group, the higher the rate of absence. Heavy manufacturing industries have high absence rates, but the sector with the highest absence rate is 'health and social services'. Absence rates are higher with higher usual hours of work. The more generous is the system, the higher the frequency and duration of absenteeism. Lastly, there is an significant effect of the unemployment rate and the business cycle in general on work-absence behavior.



## Internet Sites

Social Insurance. [www.rfv.se](http://www.rfv.se)

OECD. [www.oecd.org](http://www.oecd.org)

## Selected References

Allen, S. G. (1981a) "An Empirical Model of Work Attendance", *Review of Economics and Statistics*, Volume 63, pp. 77-87.

Allen, S. G. (1981b) "Compensation, Safety, and Absenteeism: Evidence from the Paper Industry", *Industrial and Labor Relations Review*, Volume 34, pp. 207-218.

Alexanderson, K. (1998) "Sickness Absence: A Review of Performed Studies with Focused on Levels of Exposures and Theories Utilized", *Scandinavian Journal of Social Medicine*, Volume 4, pp. 241-249.

Andrén, D. (2001) *Work, Sickness, Earnings, and Early Exits from the Labor Market: An Empirical Analysis using Swedish Longitudinal Data*. Ph.D. Thesis, Department of Economics, Göteborg University.

Arrow, K.J. (1974) *Essays in the Theory of Risk-Bearing*. Amsterdam: North-Holland.

Barmby, T.A.; M.G. Ercolani; and J.G. Treble. (2002) "Sickness Absence: An International Comparison", *Economic Journal*, Volume 112, pp. 315-331.

Barmby, T.A., C.D. Orme and J.G. Treble (1991) "Worker Absenteeism: An Analysis using Microdata", *Economic Journal*, Volume 101, pp 214-229.

Barmby, T., C.D. Orme and J.G. Treble (1996) "Worker Absence Histories: A Panel Data Study", *Labour Economics* Volume 2, pp. 53-65.

Bäckman, O. (1998) *Longitudinal Studies on Sickness Absence in Sweden*. Ph.D. Thesis,

Stockholm: Swedish Institute for Social Research.

Brown, S. and J.G. Sessions. (1996) "The Economics of Absence: Theory and Evidence", *Journal of Economic Surveys*, Volume 10, Number 1, pp. 23-31.

Henrekson, M. and M. Persson. (2002), "The Effects on Sick Leave of Changes in the Sickness Insurance System", *Journal of Labor Economics*, Volume 22, pp. 333-344.

Hensing, G., K. Alexanderson, P. Allebeck and P. Bjurulf. (1998) "How to Measure Sickness Absence? Literature Review and Suggestion of Five Basic Measures", *Scandinavian Journal of Social Medicine*, Volume 2, 133-144.

Johansson, P. and M. Palme. (1996) "Do Economic Incentives Affect Worker Absence? Empirical Evidence Using Swedish Data", *Journal of Public Economics*, Volume 59, pp. 195-218.

Johansson, P. and M. Palme (2002) "Assessing the Effect of Public Policy on Worker Absenteeism", *Journal of Human Resources*, Volume 37, pp. 381-409.

Laffont, J.-J. (1989) *The Economics of Uncertainty and Information*. Cambridge: The MIT Press.

Lantto, K. (1991) *Optimal Deterrents to Malingering*. Ph.D. Thesis, Department of Economics, Stockholm University.

Lantto, K. and E. Lindblom.(1987) "Är Arbetslösheten Hälsosam?" (Is Unemployment Healthy?), *Ekonomisk Debatt*, Volume 15, pp. 333-336.

Paringer, L. (1983) "Women and Absenteeism: Health or Economics?", *American Economic Review*, Volume 73, Number 2, Papers and Proceedings of the Ninety-Fifth Annual Meeting of the American Economic Association, pp. 123-127.

- Weiss, A. (1985) "Absenteeism and Wages",  
*Economics Letters*, Volume 19, pp. 277-279.
- Vistnes, J. P. (1997) "Gender Differences in Days Lost from Work due to Illness",  
*Industrial and Labor Relations Review*, Volume 50, pp. 304-323.

*Daniela Andrén*  
*Department of Economics*  
*Göteborg University*  
*Göteborg, Sweden*  
*Daniela.Andren@economics.gu.se*

## Taxation Compliance Costs

*Jeff Pope*

### **Introduction**

Compliance costs simply expressed are those costs that taxpayers incur as a result of meeting their taxation obligations over and above the payment of the tax itself. The term compliance costs of taxation is now widely recognised throughout many countries in the world, particularly where research studies have taken place. This understanding and acceptance of the term has not occurred overnight, or been easily achieved. Tax authorities in several countries in the world, including the UK and New Zealand, now include compliance cost assessments (variously designated) when assessing new taxation policies or amendments. Tax compliance costs are higher than, and typically a multiple of, administrative costs for nearly all taxes, based on worldwide studies to date.

There are usually, but not always, two types of tax compliance costs estimates cited in most studies: namely, gross and net (recurrent, annual or ongoing) costs. The gross costs represent the resource cost to the country and the net costs are gross costs minus offsets, such as tax deductibility and cash flow benefits. Gross and net compliance costs are also known as social compliance costs (gross) and taxpayer compliance costs (net) respectively, terms first used in Australia by Evans *et al* (1997) although most other researchers prefer the original terms gross and net as used by Sandford (eg 1973) and others. Estimates may be presented in disaggregate, i.e., at the individual or business level, and, where tax authority data permits grossing-up (see Pope 2002: 277-278), at the aggregate (country) level.

Psychological costs and managerial benefits are also recognized, but little progress has been made to date on their estimation. The literature also distinguishes between start-up, temporary and recurrent costs (see Sandford *et al* 1989: 16-18). More recent valuable conceptual and theoretical reviews, including the relationship with tax simplicity, are by Tran-Nam *et al* (2000) and Tran-Nam (2001).

Administrative costs of taxation simply expressed are those costs incurred by the government in operating the tax system. These comprise the resources employed by the government 'in imposing, assessing, levying, collecting and adjudicating on taxes' (Sandford 1973:4). Administrative costs are a form of government expenditure that delivers no direct benefits; any reduction would lead to a transfer of resources of more direct public benefit. They are typically around one to two per cent of tax revenue (or yield) for major taxes in the richest countries.

Administrative plus compliance costs (of taxation) are known as the operating costs of taxation. They can be positively or negatively correlated. For example, effective tax simplification lowers both. On the other hand, where the tax authority adopts policies to shift costs from the public to the private sector, administrative costs would be expected to fall relatively and tax compliance costs rise. A worldwide example is that of self-assessment, whereby taxpayers have to calculate their own tax liability rather than have the tax authority doing it for them. Of course self-assessment takes different forms in different countries, and is subject to tax authority review and possible audit. Tax administrative costs are generally much

lower and less contentious than compliance costs.

### **Key Findings Worldwide**

The compliance cost studies undertaken to date in various countries throughout the world nearly all demonstrate three key findings. First, compliance costs are high, whether measured in absolute money terms, as a percentage of tax paid, as a percentage of GDP or by comparison to administrative costs (the costs to the revenue authorities). Simply expressed, the most robust estimates of the compliance costs of the tax system are around one per cent of GDP in the UK and two per cent of GDP in Australia (Evans *et al* 1997:83). The most reliable and comprehensive national aggregate data on tax compliance costs emanates from the UK, Australia, Netherlands, USA and New Zealand, although the last of these is for business taxes only.

In 1994-95 prices, the gross (or social) compliance costs minus cash flow benefits in the UK and Australia were estimated at A\$7.7 billion and A\$8.8 billion respectively. This was equivalent to 2.8 per cent and 10.0 per cent of tax revenue respectively. As a percentage of GDP, the estimates represent 0.86 per cent and 1.94 per cent respectively (*ibid.*).

In the Netherlands, Allers (1994) arguably has undertaken the most comprehensive study in the world as it also included administrative and tax-benefit programmes. The operating costs of taxes and levies in the Netherlands were estimated at 11 billion Guilders in 1990, equivalent to 2.1 per cent of GDP. The estimate for the UK was 1.5 per cent with a like-for-like Netherlands comparison being 1.9 per cent (Allers 1994:187).

The operating costs of the tax-benefit system in the Netherlands were estimated at 15.3 billion Guilders in 1990, equivalent to 3.0 per cent of GDP. Allers (*ibid.*) considered that, allowing for programmes that could not be quantified or were left out of his study, the 'total annual costs of the Dutch tax-benefit system ... might amount to 20 billion Guilders ... [and] [i]ts size compared to that of the wood, furniture and building materials industry' whose production was around 16 billion Guilders in 1990.

For New Zealand, the total costs of business taxes were estimated at NZ\$1.9 billion in 1990-91 (Sandford & Hasseldine 1992:110). This was equivalent to 2.5 per cent of GDP, and over five times the total administrative costs of the Inland Revenue Department covering all taxes not just business taxes (breakdown by tax unavailable) (*ibid.*).

The costs of complying with the Federal Income Tax system in the USA have been and are continuing to rise, based on estimates and projections from 1990 to 2015 (Moody *et al* 2005). In 2005 aggregate compliance costs have been estimated at around US\$265 billion, or 22 per cent of tax revenue; business accounts for 56 per cent of this total, individuals 42 per cent and non-profit organisations for around two per cent. Put in more graphic and comparable terms, this US\$265 billion tax compliance costs figure is higher than the revenue of Walmart (US\$259 billion), America's largest company (*ibid.*:9).

Second, compliance costs are very regressive, especially for Value-Added Tax (VAT), known as the Goods and Services Tax (GST) in some countries. The level of compliance costs are a major concern to small business, unless exemptions or special arrangements for compliance exist.

Some of the most reliable data, although dated, is from the UK. For the year 1986-87, the smallest businesses (with a taxable annual turnover of less than 20,500 pounds) incurred VAT compliance costs of 1.94 per cent of (mean) taxable turnover compared with the largest businesses (taxable annual turnover of over 10 million pounds) of just 0.003 per cent (Sandford *et al* 1989:116). In other words, the smallest businesses incurred compliance costs of around 650 times higher than those large businesses than could achieve economies of scale and other benefits of size in meeting their VAT obligations. The overall weighted mean VAT compliance cost was 0.69 per cent of taxable turnover, or just under 7 pounds for every 1,000 pounds worth of goods sold. Businesses with a turnover of over 50,000 pounds pa generally incurred less than this figure.

In New Zealand, for the year 1990-91, mean GST compliance costs for the smallest businesses (with a turnover less than NZ\$30,000 pa) were 2.7% of turnover, compared with 0.005% for the largest businesses (with an annual turnover over NZ\$50 million) (Hasseldine 1995:135), a factor of 540 times greater.

Similar results confirming the importance of high regressivity of compliance costs for most taxes have been found in nearly all studies to date, including major studies in Australia (Evans *et al* 1997), Netherlands (Allers 1994) and USA (Slemrod & Venkatesh 2002). In developing countries, that generally have lower compliance costs, the regressive effect is less marked (Ariff *et al* 1995).

Third, research into compliance costs and ensuing publicity 'puts compliance costs on the political agenda'. They have played a significant role in the development by governments of the

evaluation of the compliance costs of tax regulations; also they have influenced the introduction of Taxpayers' Charters which, amongst other things, stress the duty of the tax authorities to minimise compliance costs subject to other objectives. Tax Impact Statements are nowadays generally required for any major new tax legislation in many of the world's richer countries. The worldwide development of both Taxpayers' Charters and Tax Impact Statements since the 1980s demonstrates that research into tax compliance costs has fed through and impacted upon many governments' tax policies.

The study by Evans and Walpole (1999) into Tax Impact Statements in OECD countries reviews the progress made up until the late 1990s in governments' attempts to monitor and control tax compliance costs, policy differences between OECD countries and the ongoing difficulties they face.

### **Early Recognition by Adam Smith and Subsequent Neglect**

Adam Smith (1776) first recognised the importance of the compliance costs of taxation in his four famous maxims or canons of taxation, summarised as the principles of equity, certainty, convenience and economy. It is not often appreciated that two of the four maxims are concerned wholly with compliance costs, whilst a third includes compliance costs and contains a vivid statement of the psychological costs of tax compliance. In the decades that followed Smith's inspiring work, economists placed great emphasis on the equity and efficiency aspects of taxation, particularly the deadweight loss aspects, to the almost total exclusion and neglect of the compliance costs. Even today, most public economics and finance

texts reflect this attitude, although more recent texts, and now journals, are starting to recognise the importance of compliance costs (see Tran-Nam & Evans 2002:402).

### **Early Attempts at Measurement**

The ‘pioneering expedition into this unexplored territory’ was undertaken in the USA by Professor R.M. Haig (1935), with subsequent work by J.W. Martin (1944), who recognized the fundamental characteristic that ‘there is evidence of considerable (compliance) cost which does not vary significantly in [relation to the] size of the tax bill’. Martin also stressed the difficulty of what in economics is termed a joint cost, namely identifying compliance costs accurately and separating them from ordinary business expenses and planning costs. Other notable studies in the USA included those by Yocum (1961), Muller (1963), and Wicks and Killworth (1967). The first Canadian study was by Bryden (1961).

These early North American studies are characterised by an emphasis on business compliance costs. They established a number of key characteristics, particularly that compliance costs are not directly proportional to tax liability or taxable income and that they are regressive in nature - economies of scale are likely to occur. For countries with a federal system of government, they also established that businesses that operate in multiple states are associated with high costs, particularly where states use different definitions of the tax base. The important trade-off between administrative and compliance costs was also emphasized.

About the time interest in compliance cost research seemed to be waning in North America it was beginning in Europe. The first European study was in West Germany

by Strumpel (1966), who quantified the compliance costs of small businesses and recognised their psychological burden, albeit theoretically.

### **Pioneering Work of Cedric Sandford**

The early 1970s saw the beginnings of a series of compliance cost studies initiated by Cedric Sandford at the Centre for Fiscal Studies, University of Bath, UK. This culminated in the seminal publication by Sandford et al (1989)—an essential initial reference for anyone new to the field. It includes a history of the topic, including a summary of previous studies up to 1989, definitions and concepts, and methodology of measurement, essentially a questionnaire survey technique. The major part of the book analyses and discusses the major studies undertaken by Sandford *et al* on a tax-by-tax basis. The final part estimates the administrative and compliance costs in the UK for the year 1986/87, analysing their size, distribution and effective incidence. Three measures of the importance of compliance costs were used, namely absolute money terms, as a percentage of tax revenue and as a percentage of GDP. This precedent has been followed by nearly every other study where the overall magnitude of compliance costs is estimated. The book concludes by discussing the policy implications of the research.

Sandford had the support and cooperation of the Inland Revenue and Customs and Excise in his research. The United Kingdom governments have led the world in requiring the revenue departments (as part of a more general deregulatory policy) to produce compliance cost assessments for all tax changes affecting business. Since 1986 when the policy was

first established the methodology has been refined and improved.

Sandford's name is synonymous with tax compliance costs, with many studies throughout the world adopting what may be termed the 'Sandford methodology'. Sandford has received many tributes for his contribution to the field, particularly Evans, Pope and Hasseldine (2001), *Australian Tax Forum* (2002).

### **Recent Studies Worldwide**

In the mid-1980s Slemrod made significant advances to the estimation of compliance costs in the USA, particularly for personal taxpayers. Slemrod and Sorum (1984), summarized in Blumenthal and Slemrod (1995a), estimated the total resource cost of the personal income tax system for the year 1982 based on a sample drawn from Minnesota taxpayers.

Slemrod later investigated the impact of the Tax Reform Act of 1986, finding that the Act did not lower the costs of compliance, as was the intention of policy makers (Blumenthal & Slemrod 1995a). A significant development by Slemrod in this field was the much greater use of regression analysis to identify significant variables in contributing to the level of, and changes in, compliance costs.

Whilst countries such as the UK and Australia seem particularly concerned with the effect of compliance costs on small business, US policy makers were more interested in the compliance costs of the large corporations (Slemrod & Blumenthal 1996). Other important US studies include the cost of itemising deductions in the US income tax (Pitt and Slemrod 1988), the compliance costs of the federal income tax system (Arthur D. Little 1988, for the US Treasury, and estimated only in hours) and

foreign-source income (Blumenthal and Slemrod 1995b).

In Canada, major studies include Vaillancourt (1989), who used face-to-face interviews and a national commercial polling organisation (omnibus survey), Erard (1997), Plamondon (1991; 1995; 1997).

More recent major studies not already discussed include a study on the tax compliance costs of employers' PAYE and National Insurance in the UK (Collard *et al* 1998) and the development of a compliance burden model in the USA (Stavrianos and Greenland 2002).

An issue of concern for federal systems of government is whether a single tax collection agency would be likely to lead to lower administrative and compliance costs compared with those arising from a combination of federal and provincial, territorial and/or state tax administrations. Canadian research (Public Policy Forum 1997; Plamondon and Zussman 1998) found that such a single agency would be beneficial and produce significant cost savings.

Whilst the topic of compliance costs has developed most rapidly in English-speaking countries, particularly in the last two decades, studies in other countries, particularly Europe, should not be ignored. Important compliance cost studies have taken place in the Netherlands (Allers, 1994), Spain (Diaz & Delgado 1995; Delgado *et al* 2001) and Sweden (Malmer 1995). The Allers (1994) research is particularly noteworthy as it comprised an investigation of transfers between the public and private sectors in the Netherlands, and is probably the single most comprehensive study to date anywhere in the world.

Australia and New Zealand have been at the forefront of tax compliance costs research since the late 1980s, with a series of five major Australian studies by Pope *et al* (summarized in Pope 1995; see also Pope *et al* 1993), Evans *et al* (1996 and 1997), Pope *et al* (2003), and, in New Zealand, Sandford and Hasseldine (1992).

Research in the Asia-Pacific region comprises a series of studies on the compliance costs of corporate income taxation in Singapore, Malaysia and Hong Kong (Ariff & Pope 2002). Up until this work tax compliance costs could best be described as a 'rich, mainly OECD, country topic'. Two studies have also been published in India (Chattopadhyay *et al* 2002a,b).

Studies in other parts of the world are few and far between, but do include those in Africa, on excise duties (Shekidele 1999); Mauritius, on VAT (Pillai 2000); Brazil, on public companies (Bertolucci 2002); and Croatia, on personal income tax (Blazic 2004).

The effect of Tax Impact Statements throughout the OECD has been investigated (Evans and Walpole 1999), and the compliance costs of international business transactions and cross-border issues are now being recognized (e.g. Verwaal 2000). International organizations such as the OECD (Cordova-Novion & de Young 2001) and the European Commission (private communication, 2001) are increasingly showing interest in business tax compliance costs and related issues.

Comparative studies between different countries are rare, although Evans (2003a) did compare the operating costs of capital gains for individuals in both Australia and the UK.

Overall, tax compliance costs research has been diffuse in terms of the types of taxes covered, although any classification by tax is difficult as some studies cover a number of usually closely-related taxes. One classification of major studies from 1980 up until 2003 (Evans 2003b) found that worldwide, roughly, there were around twelve taxes solely on personal income taxes and about the same number on corporate income taxes. Eight other studies considered all federal taxes, whether business or personal. There were seven studies specifically on the Value Added Tax/Goods and Services Tax (VAT/GST) and four on sales tax and excise duties.

Other studies have covered specialist taxes or areas such as tax expenditures (Gunz *et al* 1995), petroleum revenue tax (Sandford *et al* 1989), taxes upon superannuation (Pope *et al* 2003) and road tolls (Friedman & Waldfogel 1995). Research into the tax compliance costs of businesses has been fairly evenly split between that of large and small businesses, although generally most studies cover all business sizes.

### **Methodological and Measurement Issues**

Tax compliance cost studies mainly use quantitative techniques and a deductive methodological approach, although some studies do use or include a qualitative approach. Evans (2003b:70) considers that '[t]heories develop through the study rather than being established at the outset, and the methodology is also likely to emerge as part of the research than being predetermined.'

More specifically, a variety of methodologies are used in tax compliance cost studies, including surveys (usually using questionnaires, through the medium of post, telephone, email or commercial



polling or research organisations), other interview-based techniques (eg face-to-face, with a structured survey instrument), diary and case study approaches, and estimation/simulation techniques. Some studies adopt a combination of these methodologies. The most popular methodology is that of the postal questionnaire (around two-thirds of all studies, as noted by Evans 2003b:71), followed by interview techniques (around a quarter). Studies featuring diary and case studies were the least popular.

There was a wide range in response rates to postal surveys, from less than 10% up to 50%, although few researchers would disagree with a 'norm' of around 30%, first identified by Sandford (1995:379). Studies with higher response rates often have features such as being well-funded, the support of the tax authority, support from leading professional bodies and prize incentives eg the chance to win a computer.

An important and complex measurement issue long recognized by researchers is that of joint or overhead costs, namely the allocation of business costs between tax compliance activities and normal business activities independent of the tax system (see Sandford *et al* 1989:14-16). Nearly all studies leave such an allocation implicitly to the respondent, although respondents' compliance cost estimates are usually subject to a thorough checking and verification procedure.

There is no unanimity amongst researchers on the most appropriate method for valuing individual taxpayers' time. Four major methods have been used: each individual's reported value, possibly subject to some maximum rate; the sum that taxpayers would pay to be rid of all compliance costs (a hypothetical concept which most respondents find very difficult

to handle); the before tax wage rate; and the after tax wage rate (see Pope 1995:114-118; Evans *et al* 1997:9-12).

Apart from one US study (Arthur D. Little, 1988), using hours alone has not been used by other researchers, although most studies do cite time spent in hours before converting them into a monetary estimate. In such cases the significance of the conversion measure used is then clearly apparent. Several studies have followed this route, but perhaps less explicitly than is desirable.

The US study by Moody, Warcholik and Hodge (2005:14-15) is based upon Internal Revenue Service (IRS) estimates of the time taken to complete various income tax forms (known as the paperwork burden), with time estimates then converted into monetary estimates using the appropriate hourly wage rate plus benefits (opportunity cost) figure for both individual filers and those using tax professionals. It is worth noting that from 2006 the IRS has replaced its long-standing form-by-form time estimates by a new system that gives one collective estimate of the compliance burden for all individual taxpayers (*ibid.*:16).

One of the most important methodological developments has been made by Allers (1994) of the Netherlands. Briefly, Allers developed a 'post-card method' using a single question in order to test for non-response bias, in reply to critics of compliance cost postal questionnaire surveys, such as Tait (1988). He found that estimates were biased downward by non-response rather than upwards as theoretically argued by the critics. The Allers method has subsequently been used successfully in other studies including an Australian study by Evans *et al* (1996:37; 1997).

The research, community and government interest in compliance costs of taxation in many of the world's leading and most prosperous countries in the 1980s and 1990s was largely unco-ordinated, with each country's research developing independently. The topic was sufficiently small and specialised for researchers to monitor each study (eg Sandford et al 1989). A significant advance to knowledge and the literature in the field was instigated by Sandford, who organised a closed conference at Oxford on compliance cost measurement and policy issues in 1994. Leading academic researchers, together with senior tax administration and government officials, from throughout the world reviewed the methodologies of the major studies and assessed their value to policy-makers (Sandford 1995).

The two major survey methodologies used by compliance cost researchers to date are evaluated, namely large-scale surveys of taxpayers (particularly the postal questionnaire survey technique, sampling methods, and grossing-up procedures) and depth surveys of taxpayers and tax professionals. Sandford (1995:375-401) critically evaluates the methodologies and suggests ways of improving them, as well as analysing many of the key questions raised by critics, such as non-response bias, sampling difficulties and the reliability of response estimates. Perhaps the greatest value of this ex post review was the interaction of policy-makers with researchers, the outcome being a much clearer picture of the next steps in the development of the topic (Sandford 1995: 402-413). A second tax compliance costs research conference was held in Sydney in 2000 that followed a similar format to the 1994 Oxford conference, updating methodological developments and

estimation findings from throughout the world (Evans, Pope and Hasseldine 2001).

### **Current Major Issues**

A number of key issues in the tax compliance costs field may be identified from a global perspective. Such issues are more widely based and focus on government tax policy, and aspects between tax policy and tax compliance costs research.

One particular focus is on small business tax compliance costs, where nearly all the evidence to date clearly shows that small business bears a heavy disproportionate burden in comparison to large business when compliance costs are expressed as a percentage of annual turnover or tax collected. There is a common perception, certainly in the USA and probably in many other countries that small business owners engage in more tax evasion than larger businesses or wage earners. Slemrod (2004:90) found that there is 'substantial evidence' in the USA of the extent of such evasion compared with such income sources as wages, salaries, interest and dividends. Slemrod (2004:94) then suggests that, roughly and subject to several caveats, particularly that estimates are based on averages, small business tax compliance costs are offset by tax evasion. In the tax compliance costs field this is an important finding that demonstrates the important and the research-neglected relationship between tax compliance costs and compliance.

The other major issue in the field that ebbs and wanes, and that many researchers and policy makers often like to ignore, concerns the degree of political recognition and respect for the field. In short, tax compliance costs are politically very sensitive and the 'poor relation' when

compared to other tax policy issues, particularly compliance and equity. In other words some governments give the simplicity and efficiency aims of taxation policy 'lip service' yet effectively do all they can to minimize research (eg through failure to support either financially or non-financially) and/or publicity on the topic (eg failure to publish commissioned research). There are some fears that such policies may be emerging in some of the leading countries in the field, particularly Australia and the UK.

Pope (1993:71-73) identified a number of possible stages in the awareness of tax compliance costs. Initially there would be neglect, then recognition by tax professionals and advisers, followed by quantification (usually by academic researchers), policy recognition, effective policy measures leading to lower compliance costs, and finally continual monitoring of compliance costs.

Evans (2003b:73) recognizes this progression, and considers that many of the developed countries that have undertaken compliance costs research have certainly reached the policy recognition phase. Some even pay lip-service to compliance costs when formulating tax policy. However, overall, Evans is rather pessimistic, stating that "there is still a long way to go before any of the countries can truly claim that its tax laws are designed with a clear focus on the implications that design will have on the operating costs [ie administrative and compliance costs] of the tax system." This almost certainly succinctly represents the current view of many, and probably the vast majority of, researchers in the field.

### **Conclusions and Further Research**

Research into the compliance costs of taxation, a topic practically unheard of in

the early 1980s, seems to have reached a critical position in terms of its historical development and policy impact. Pope and Sandford (1999) had earlier been much more optimistic, arguing that the field seemed to be going from strength to strength in many of the leading and richer countries throughout the world. Further, governments in these countries were not only becoming increasingly aware of the political importance of compliance costs, but specifically incorporating compliance costs into their tax policy decision-making processes.

The current view in terms of policy application is more cautious and possibly even pessimistic, as illustrated by Evans above, although it must be emphasized that tax compliance costs develop, advance and are incorporated into government tax policy at different rates depending upon the particular country in question. For example, New Zealand seems to have been particularly successful in its research studies and ensuing policies; eg Sandford and Hasseldine 1992; Victoria University of Wellington 2006.

What is encouraging is the number of countries from 'outside the Anglo-Saxon fold' more recently entering into compliance costs research eg Spain, India, Croatia, or strengthening the little research they have to date eg Malaysia. In addition, the importance of compliance costs vis-a-vis deadweight losses is beginning to be recognised within the economics profession as a whole.

Areas of further research focus include or are likely to include: psychological costs eg Woellner, R. *et al* (2001); start-up costs eg Rametse and Pope (2002); managerial benefits of tax compliance; the effects of information technology eg Hasseldine and Hansford (2002), Goolsbee (2004); the use

of pre-populated returns (eg Highfield 2006); international comparisons using the same survey instrument by international or regional bodies eg OECD (Cordova-Novion & de Young 2001); in many developing countries, design of effective tax regimes that could keep their existing lower tax compliance costs while modernizing the tax system and ensuing administration.

*Acknowledgment:* The author is grateful for the comments of Professor Mohamed Ariff and the anonymous referees on an earlier draft.

### **Selected References**

Allers, M.A. (1994) *Administrative and Compliance Costs of Taxation and Public Transfers in the Netherlands*. Groningen: Wolters-Noordhoff.

Ariff, M. and J. Pope. (2002) *Taxation and Compliance Costs in Asia Pacific Economies*. Sintok, Malaysia: Universiti Utara Malaysia Press.

Ariff, M., A. Loh, Z. Ismail and M. Shamsheer. (1995) "Compliance Costs of Corporate Income Taxation in Malaysia, 1995", *Pacific Accounting Review*, Volume 14 Number 3, pp 134-176.

*Australian Tax Forum*. (2002) "Tribute Edition to Cedric Sandford", Volume 17 Number 4.

Bertolucci, A. (2002) "The Compliance Costs of taxation in Brazil: A Survey of Costs in Brazilian Public Companies", Faculdade de Economia, Administracao E Contabilidade, University of Sao Paulo, unpublished paper.

Blazic, H. (2004) "Personal Income Tax Compliance Costs at an Individual Level in Croatia", *Australian Tax Forum*, Volume 19, Number 3, pp325-352.

Blumenthal, M. and J. Slemrod. (1995a) "Recent Tax Compliance Research in the United States", in Sandford (1995) (Editor), pp142-172.

Blumenthal, M. and J. Slemrod. (1995b) "The Compliance Cost of Taxing Foreign-Source Income: Its Magnitude, Determinants, and Policy Implications", *International Tax and Public Finance*, Volume 2, May, pp. 37-53.

Bryden, M.H. (1961) *The Cost of Tax Compliance*. Canadian Tax Foundation Paper, No. 25.

Chattopadhyay, Saumen and Arindam Das-Gupta. (2002a) *The Compliance Cost of the Personal Income Tax and its Determinants*. New Delhi: National Institute of Public Finance and Policy.

Chattopadhyay, Saumen and Arindam Das-Gupta. (2002b) *The Income Tax Compliance Cost of Indian Corporations*. New Delhi: National Institute of Public Finance and Policy.

Collard, D., S. Green, M. Godwin and L. Maskell. (1998) "The Tax Compliance Costs for Employers of PAYE and National Insurance in 1995-96", *Inland Revenue Economics Paper No. 3*. London: Inland Revenue.

Cordova-Novion, C. and C. de Young. (2001) "The OECD PUMA Multi Country Business Survey—Benchmarking the Regulatory and Business Environment", in C. Evans, J. Pope and J. Hasseldine (2001) *Taxation Compliance Costs: A Festschrift for Cedric Sandford*. Sydney: Prospect Media Pty Ltd, pp. 205-228.

Diaz, C. and M.L. Delgado. (1995) "The Compliance Costs of Personal Income Tax in Spain", in Sandford (1995) (Editor), pp. 210-225.

Delgado, M.L., J. Salinas-Jimenez and J.F. Sanz. (2001) "Hidden Tax Burden of the

- Personal Income Tax in Spain: Evidence from the Recent Tax Reform in Spain”, *Australian Tax Forum*, Volume 16 Number 4, pp. 463-482.
- Erard, B. (1997) “The Income Tax Compliance Burden on Canadian Big Business”, *Working Paper 97/2*. Ottawa: Technical Committee on Business Taxation, Department of Finance.
- Evans C. (2003a) *The Operating Costs of Taxing the Capital Gains of Individuals: A Comparative Study of Australia and the UK, with Particular Reference to the Compliance Costs of Certain Tax Design Features*. PhD Thesis. Sydney: University of New South Wales.
- Evans C. (2003b) “Studying the Studies: An overview of recent research into tax operating costs”, *eJournal of Tax Research*, Volume 1 Number 1, pp 64-92.
- Evans, C. and M. Walpole. (1999) *Compliance Cost Control: A Review of Tax Impact Statements in the OECD*. Sydney: Australian Tax Research Foundation.
- Evans C., J. Pope and J. Hasseldine. (2001) (Editors) *Taxation Compliance Costs: A Festschrift for Cedric Sandford*. Sydney: Prospect Media Pty Ltd.
- Evans, C., K. Ritchie, Binh Tran-Nam and M. Walpole. (1996) *Costs of Taxpayer Compliance, Final Report*. Sydney: Prepared for the Revenue Analysis Branch, Australian Taxation Office, University of NSW (ATAX Programme).
- Evans, C., K. Ritchie, Binh Tran-Nam and M. Walpole. (1997) *Taxpayer Costs of Compliance*. Canberra: Australian Taxation Office, AGPS.
- Friedman D. and J. Waldfogel. (1995) “The Administrative and Compliance Costs of Manual Highway Collection: Evidence from Massachusetts and New Jersey”, *National Tax Journal*, Volume 48, Number 2, pp. 217-228.
- Goolsbee A. (2004) “The TurboTax Revolution: Can Technology Solve Tax Complexity”, in Aaron H. and Slemrod J. (Editors), *The Crisis in Tax Administration*. Washington DC: Brookings Institute, pp.124-147.
- Haig, R.M. (1935) “The Cost to Business Concerns of Compliance with Tax Laws”, *Management Review*, pp. 232-333.
- Hasseldine, J. (1995) “Compliance Costs of Business Taxes in New Zealand”, in Sandford (1995) (Editor), pp. 126-141.
- Hasseldine, J. and A. Hansford. (2002) “The Compliance Burden of VAT: Further Evidence from the UK”, *Australian Tax Forum*, Volume 17 No. 4, 369-387.
- Highfield, R. (2006) “Pre-Populated Income Tax Returns: The Next ‘Big Thing’ in Reform of the Administration of Australia’s Personal Income Tax System?” Paper presented at the 7<sup>th</sup> International Tax Administration Conference, Sydney, 19-21 April.
- Little, Arthur D. (1988) Corporation, *Development of Methodology for Estimating Taxpayer Paperwork Burden, Final Report to Department of Treasury*. Washington DC: Internal Revenue Service.
- Malmer, H. (1995) “The Swedish Tax Reform in 1990-91 and Tax Compliance Costs in Sweden”, in Sandford (1995) (Editor), pp. 226-262.
- Martin, J.W. (1944) “Costs of Tax Administration, Examples of Compliance Expenses”, *Bulletin of National Tax Association*, April, pp. 194-205.
- Moody, J.S., W.P. Warcholik and S.A. Hodge. (2005) “The Rising Costs of

- Complying with the Federal Income Tax”, *Tax Foundation Report No. 138*. [www.taxfoundation.org/publications](http://www.taxfoundation.org/publications)
- Muller, F.J. (1963) *The Burden of Compliance*. Seattle: Seattle Bureau of Business Research.
- Pillai, K.V. (2000) “The Compliance Costs of VAT in the Hotel Industry in Mauritius”, Social Sciences and Humanities and Law & Management, University of Mauritius, *Research Journal*, Volume 3 No. 7, pp 52-66.
- Pitt, M. and J. Slemrod (1988) *The Compliance Cost of Itemising Deductions: Evidence from Individual Tax Returns*, mimeo.
- Plamondon, R., (1991) *GST Compliance Costs for Small Business in Canada*. Ottawa: Department of Finance, Government of Canada (updated by Plamondon and Associates Inc., 1993).
- Plamondon, R. (1995) *GST: The Quick Method of Accounting*. Ottawa: A Report for Revenue Canada and Customs and Excise, Government of Canada, Plamondon and Associates Inc..
- Plamondon, R. (1997) *Business Compliance Costs for Federal and Provincial Payroll Taxes*. Ottawa: Plamondon and Associates Inc.
- Plamondon, R. and D. Zussman. (1998) “The Compliance Costs of Canada’s Major Tax Systems and the Impact of Single Administration”, *Canadian Tax Journal*, Volume 46, Number 4, pp761-785.
- Pope, J. (1993) “The Compliance Costs of Taxation in Australia and Tax Simplification: The Issues”, *Australian Journal of Management*, Volume 18 Number 1, pp 69-89.
- Pope, J. (1995) “The Compliance Costs of Major Taxes in Australia”, in Sandford (1995) (Editor), pp 101-125.
- Pope, J. (2002) “Administrative and Compliance Costs of International Taxation”, in A. Lymer and J. Hasseldine (Editor), *The International Taxation System*. London: Kluwer, pp. 273-289.
- Pope, J. and C. Sandford. (1999) “Tax Compliance Costs” in M. Lamb and A. Lymer (Editors), 1999, *Interdisciplinary Research in Taxation: Research Approaches and Bibliographical Survey*. London: Institute of Chartered Accountants in England and Wales, pp 85-98.
- Pope, J., R. Fayle and D.L. Chen. (1993) *The Compliance Costs of Employment-Related Taxation*. Australian Tax Research Foundation, Sydney.
- Pope, J., P. Fernandez and V. Le. (2003) *The Hidden Costs of the Superannuation Surcharge Tax*, Australian Tax Research Foundation, Sydney.
- Public Policy Forum (1997) *Cutting the Costs of Tax Collection Down to Size*, Ottawa. Toronto: Prepared for Revenue Canada; principal author: R. Plamondon of Plamondon and Assoc. Inc.
- Rametse, N. and J. Pope. (2002) “Start-up Tax Compliance Costs of the GST: Empirical Evidence from Western Australian Small Businesses”, *Australian Tax Forum*, Volume 17, Number 4, pp. 407-442.
- Sandford, C. (1973) *Hidden Costs of Taxation*. London: Institute for Fiscal Studies.
- Sandford, C. (1995) (Editor) *Tax Compliance Costs Measurement and Policy*. Bath: Fiscal Publications.
- Sandford, C. and J. Hasseldine. (1992) *The Compliance Costs of Business Taxes in New Zealand*. Wellington: Institute of Policy Studies, Victoria University of Wellington.



- Sandford, C.T., M.R. Godwin and P.J.W. Hardwick. (1989) *Administrative and Compliance Costs of Taxation*. Bath: Fiscal Publications.
- Shekidele, C. (1999) "Measuring the Compliance Costs of Taxation Excise Duties 1995-96", *The African Journal of Finance and Management*, Volume 7 Number 2, pp. 72-84.
- Slemrod, J. (2004) "Small Business and the Tax System", in H. Aaron and J. Slemrod (Editors), *The Crisis in Tax Administration*. Washington DC: Brookings Institute, pp 69-123.
- Slemrod, J. and M. Blumenthal. (1996) "The Income Tax Compliance Cost of Big Business", *Public Finance Quarterly*, Volume 24, Number 4, pp. 411-438.
- Slemrod, J. and N. Sorum. (1984) "The Compliance Cost of the US Individual Income Tax System", *National Tax Journal*, Volume 37, pp. 461-474.
- Slemrod, J. and V. Venkatesh. (2002), "The Income Tax Compliance Cost of Large and Mid-Size Businesses, A report to the IRS Large and Mid-Size Division". Michigan: Office of Tax Policy Research, University of Michigan Business School, September.
- Smith, A. (1776) *Inquiry into the Nature and Causes of the Wealth of Nations*. London: Everyman Edition, 1977.
- Stavrianos, M. and A. Greenland. (2002) "Design and Development of the Wage and Investment Compliance Burden Model". Paper presented at the *IRS Research Conference USA*: [www.irs.gov/pub/irssoi/wgincobm](http://www.irs.gov/pub/irssoi/wgincobm)
- Strumpel, B. (1966) "The Disguised Tax Burden", *National Tax Journal*, Volume 19, Number 1, pp. 70-77.
- Tait, A.A. (1988) *Value Added Tax: International Practice and Problems*. Washington DC: International Monetary Fund.
- Tran-Nam, B. (2001) "Tax Compliance Costs Methodology—A Research Agenda for the Future", in C. Evans, J. Pope and J. Hasseldine (Editors), *Taxation Compliance Costs: A Festschrift for Cedric Sandford*. Sydney: Prospect Media Pty Ltd, pp. 51-68.
- Tran-Nam, B. and C. Evans. (2002) "The Impact of Cedric Sandford on the Discipline of Tax Compliance Costs", *Australian Tax Forum*, Volume 17 Number 4, pp. 389-405.
- Tran-Nam, B., C. Evans, K. Ritchie and M. Walpole. (2000) "Tax Compliance Costs: Research Methodology and Empirical Evidence from Australia", *National Tax Journal*, Volume 53, pp. 229-252.
- Vaillancourt, F. (1989) *The Administrative and Compliance Costs of the Personal Income Tax and Payroll Tax System in Canada*. Toronto: Canadian Tax Foundation. Summarized in F. Vaillancourt (1995), "The Compliance Costs of Individuals in Canada: Personal Income Tax and Payroll Taxes", in Sandford (1995) (Editor), pp. 196-209.
- Verwaal, E. (2000) *Compliance Costs of International Business Transactions in the European Union*. Totterdam: Faculty of Economics, Erasmus University.
- Victoria University of Wellington. (2006) "Twenty Years of GST: the Best Path Forward", *International Conference* 16-18 November. Wellington: Centre for Accounting, Governance and Taxation.
- Wicks, J. and Killworth N. (1967) "Administrative and Compliance Costs of State and Local Taxes", *National Tax Journal*, Volume 20, Number 33, pp. 309-315.
- Woellner, R. et al (2001) "Taxation or Vexation—Measuring the Psychological

Costs of Taxation”, in C. Evans, J. Pope and J. Hasseldine (Editors), *Taxation Compliance Costs: A Festschrift for Cedric Sandford*. Sydney: Prospect Media Pty Ltd, pp. 35-49.

Yocum, J.C. (1961) *Retailers' Costs of Sales Tax Collection in Ohio*. Ohio State University Bureau of Business Research.

*Jeff Pope*  
*Tax Policy Research Unit*  
*Department of Economics*  
*Curtin University*  
*Perth, Australia*  
*jeff.pope@cbs.curtin.edu.au*



## Taxes-Drive-Money Approach

*Sergio Rossi*

### **Introduction**

The taxes-drive-money (TDM) approach is an alternative view of money with respect to the orthodox seigniorage approach. It is based on the idea that both the existence and validity of money are strictly related to the power of the state to impose various forms of tax liabilities onto the general public. Wray (2003a:89) labels this view as a sovereignty approach to money, “because it links the state’s ability to issue a currency denominated in the unit of account it has chosen ... to a fundamental power that is directly associated with sovereign nations.”

### **Historical Roots and Evidence**

The TDM approach is strongly rooted into the “state theory of money” view put forward by several chartalists in the history of monetary thought (see Smith 1776; Knapp 1905; Keynes 1930). It has been revived by some neo-chartalists at the turn of the new millennium, and may be considered today as a theoretical synthesis between the exogenous and endogenous money views (see Goodhart 1998; Wray 1998, 2003a,b, 2004; Bell 2000, 2001; Bell & Nell 2003).

For the metallist school, as is well known, money is a creature of the markets, developed to solve the so-called double coincidence of wants problem. By contrast, (neo)chartalists explain that money is a creature of the state, since the latter is entitled by law “to declare what thing should answer as money to the current money of account” (Keynes 1930:4). In light of this principle, TDM proponents argue that fiat money is a liability of the

state, which the latter agrees to accept at state pay offices and that therefore circulates widely in the country (see Wray 2003a:89–90). Elaborating on this point, they argue further that the emission, circulation, and validation of money depend on the state’s willingness to accept its own liabilities (IOUs) as a means of payment for extinguishing any sort of debt obligations that its citizens may have to it (not only taxes, but also fees, fines, duties, tithes, interests, user charges, and so on; see Lerner 1947).

The role of the state in both ancient and modern monetary economies is thus stronger in the TDM view than in chartalism. In fact, according to the TDM view, this role is now more important than it used to be in the past, that is, in ancient societies. Proponents of this (neo-chartalist) perspective indeed emphasise that the sophistication of the penal system as well as of many fiscal instruments elicited the need for the state to define a standardised unit of account in order for it and its population to record and finally settle any debt obligations.

To substantiate their approach on factual as well as historical grounds, adherents to the TDM view indicate as empirical evidence that the state has been keeping track of its outlays and receipts with various accounting methods as time went by, in particular with elementary bookkeeping systems and/or fiscal notes of different sorts, some fragments of which were indeed found in centres of state power (like palaces and temples). This shows that money does not need to be reified into precious metal to be a means of payment: it would be enough that a government keeps a double-entry book by means of which its economic transactions are recorded and settled with merely a book-entry device. In

line with chartalism, TDM proponents further argue that coins represent only one possible form of book-entry money and, indeed, have been used rather widely owing to the state's monopoly on the mints, and also in order to make counterfeiting more difficult. In fact, in the TDM view state money is fiat money, in the form of token money but even more so in the form of a double-entry bookkeeping in a state's ledger.

In this regard, Courbis et al. (1991:323–5) provide strong arguments according to which paper was discovered very early in history but only became a form of money much later on: paper had been for a long historical period the material used by human beings in order for them to keep accounting books in what were once called “*banche di scritta*” (that is to say, accounting banks) in Italy, and particularly in fourteenth-century Venice. Often, these banks were public or state banks with a rudimentary but efficient book-entry payments system.

This system was indeed developed during the fourteenth century, probably in Genoa and Florence (Copeland 1981:254, fn. 6), even if it was already used in Egypt about 300 or 200 B.C. As Rostovtzeff (1941:404–6) reports, a second-century papyrus record found during some archaeological work contains fragments of the daily payments recorded by a small state bank in the province of Heracleopolitan: “In many instances [...] the payments were effected by transfer from one account to another without money passing.” The same system was also in use in Greece at least by the fourth century B.C. (Copeland 1981:249): when a debtor settled his debt obligation, his account in the state bank was debited and the creditor's account credited, with neither

coins nor notes moving around. Hence, the money creation process might be viewed as a state's record of economic transactions in its books.

### **State and Money Creation Process**

In the view of TDM proponents, the state can and does create money by a stroke of the pen, at its will, because it is in a position to spend before earning an income, that is, tax receipts. According to this view, the role of money is to keep track of government outlays on goods and services, delivered by taxpayers (citizens and private-sector businesses) in exchange for fiat (state) money, which these agents require in order for them to pay their tax obligations. To wit, fiat money is a form of credit that its issuer asks for, and obtains, from those agents giving up goods and services in exchange for it. In the words of Wray (1998:80), “[w]hen the government creates fiat money to purchase goods and services [...], this shows up on the books of the public as a credit of fiat money and a debit of goods and services sold to the government [...]. This is ‘net money creation’ because it is not offset by a private sector liability. This ‘net money’ (also called ‘outside money’) is available to pay taxes.” Indeed, state money is called “outside money,” because it enters the private sector of a country from outside of it, namely, from the public sector of this same country. Within the private sector, state money is thus considered as a net asset, because, for this sector, it is an asset to which there corresponds no liability.

In the TDM view, therefore, outside (state) money is injected into the economy through government spending on goods and services (including labour services). The state's IOUs are then deposited into the banking system by their recipients, and

this creates bank reserves that may then result in an expansion of banks' assets and liabilities. This amounts to saying that state money is exogenous and that bank money is a multiple of it, as in the orthodox money multiplier story (see Wray 1998:111).

Contrary to orthodox theory, however, the TDM approach considers that the government accounting identity has not to be written as:

$$(1) \quad G + Tr + iB \equiv T + \Delta B + M,$$

where  $G$  represents government spending,  $Tr$  are social security transfers,  $iB$  is government interest payments on bonds,  $T$  are taxes,  $\Delta B$  represents the issue of new government debt, and  $M$  stands for the money supply. In fact, in this identity the money supply would just be a source of government financing among others, and owing to today's constraints on monetary as well as fiscal policies, it would not allow a country to enhance real GDP growth through government spending: as a matter of fact, orthodox economic policies aim at reducing taxes, balancing public budgets, and limiting the growth rate of the money supply in order to guarantee price level stability.

By contrast, proponents of the TDM view depict the government accounting identity as follows:

$$(2) \quad G + Tr + iB \equiv M \equiv T + \Delta B + \Delta H,$$

where  $\Delta H$  represents the change in hoarding by private sector agents (see Kadmos & O'Hara 2000:3–5). This identity means that a government's outlays (shown in the left hand side of it) are entirely financed by the money creation process, which gives rise to a supply of

money,  $M$ , that, in a subsequent period, is partly absorbed through taxes: what remains as a money stock, in the end, depends on the amount of government budget deficit and on changes in hoarding by the population.

Hence, the normal course of events in the TDM view is for any governments to have a budget deficit, because the population needs a money stock for transaction motives. In order to provide liquidity to the economic system, therefore, the state must run a budget deficit and has to finance its expenditures by money creation. As a result, contrary to traditional analysis, fiscal policy determines the money stock and monetary policy determines interest rates via open market operations whose main object is the sale and purchase of government bonds, in order for the central bank to drain or inject reserves into the system, and thus attain the targeted policy interest rates.

As Wray (2003a:97) observes, however, “[t]his does not mean that deficit spending is always desirable, nor does it mean that government should ignore impacts of deficits or exogenously set low interest rates on domestic inflation or on exchange rates.” According to the TDM view, in fact, there are some limits – mainly institutional or linked to confidence – to the government's ability to create money without creating inflation at the same time and pace. “If the tax system breaks down, the government's fiat money can become worthless—which is manifested as ‘hyperinflation’” (Wray 1998:85). Further, the extent to which government debt-financed spending refers to production activities rather than public consumption has an impact on the real value of money as well as on the causes of inflation. As a matter of fact, (state) money created to

finance public sector consumption could give rise to some inflationary pressures on prices, whereas money created to finance a production of public goods or services has no such effect, because in this case the money–output relationship is not affected by the latter creation. In connection to this point, empirical literature on the kinds of government outlays shows that there are two opposing forces at work in the system (see, e.g., Aschauer 1989; Barro 1990; Miller & Russek 1997; Ahmed & Miller 2000): higher public sector investment might increase the level of interest rates and thus crowd out private investment; but it may also raise profitability in private sector activities and, therefore, crowd in some investment by the latter sector.

For example, in the case of transportation and communication investments, government spending often stimulates private investment, which can spur a productivity push during the building-up of the relevant infrastructure (see Fernald 1999:621). This explains the finding of Easterly and Rebelo (1993) that this kind of public investment leads to higher growth in real per capita GDP in developing countries, where these infrastructures are in their build-up stage, but not in developed countries. By way of contrast, empirical evidence shows that government spending on social security and welfare actually crowds out investment in developed as well as in developing countries. Also, “tax-financed government expenditure, in general, crowds out investment more frequently than debt-financed government expenditure” (Ahmed & Miller 2000:132), owing to the existence of liquidity constraints in the economy. This may lend some support to the TDM view of the aim and workings of fiscal and monetary policies.

## **Fiscal and Monetary Policies**

The TDM approach to government policy is a way to turn the rationale for state intervention on its head. Fiscal policy, in the sense of the state’s decision to spend and tax, has to make sure that there is enough money circulating in the economy, and in the public’s hoards. This is thus not the task of monetary policy, as is traditionally thought, which is to determine or at least influence the structure of interest rates through bond sales and purchases (see Wray 1998:34).

According to TDM proponents, bond sales reduce bank reserves, as banks pay for them with an amount of central bank money, which TDM adherents consider as state money flowing back to the treasury. In fact, as Wray (2003a:92) acknowledges, the treasury is not the unique source of reserves absorption and injection: also the central bank intervenes in open markets on a regular basis for reserve maintenance goals. Additionally, the central bank engages in a number of transactions with the government (via its treasury), which, however, Wray (2003a:92) abstracts from in his analysis, because he considers them like “accounting manoeuvres practised by husband and wife within the household”.

Now, while the TDM approach turns the rationale for both fiscal and monetary policies upside down, it still considers the government responsible for macroeconomic stabilisation: government has to guarantee full employment as well as price stability (Wray 1998). In this respect, TDM proponents advocate the Lerner (1943) principle of functional finance, instead of the sound finance doctrine propounded by the orthodoxy in terms of a balanced budget. “The central idea is that government fiscal policy, its spending and taxing, its borrowing and

repayment of loans, its issue of new money and its withdrawal of money, shall all be undertaken with an eye only to the *results* of these actions on the economy and not to any established traditional doctrine about what is sound or unsound” (Lerner 1943:39).

The basic argument in favour of functional finance is that the state must do something to guarantee a high level of employment, since private sector demand is inadequate for it owing to liquidity preference, that is to say, excessive private savings in respect of private investment. In terms of the national income equation, this situation is captured by the following expression:

$$(3) \quad G - T = (S - I) + (M - X),$$

where  $S$  represents private saving,  $I$  is private investment,  $M$  are imports, and  $X$  are exports. This expression tells us that if there is excess private savings ( $S - I > 0$ ), then the public sector must spend more than what it collects as taxes, on the assumption that foreign trade is balanced or that the economy is closed. As Sawyer (2003:882–3) puts it, “[i]f the right side of this equation would be positive at full employment, then the left side necessarily would be positive, in other words, a budget deficit would be required to sustain full employment”.

### State as Employer of Last Resort

On the ground of the TDM view, the state can and should act as an “employer of last resort” (ELR) to reach a high, if not full, employment level. According to Wray (1998), government should hire all those unemployed who are ready, willing, and able to work at a fixed minimum wage determined by it (for similar ELR or

“labour buffer stock” programmes see also Gordon 1997; Mitchell & Watts 1997; Forstater 1998; Mitchell 1998). “The government would essentially stand ready to ‘buy’ or ‘sell’ labor, offering jobs to any workers who showed up, or offering workers to any employers willing to hire workers out of the buffer stock” (Wray 2003b:106). In an economic recovery or boom, this buffer stock would shrink, contributing thus to limit wage increases, because of the fixed wage floor that the state defines by holding constant the wage rate it offers. During a recession, by contrast, buffer stock employment would rise, thereby preventing wages from falling below the rate that the government defines in its programme (see Kadmos & O’Hara 2000:9, Fig. 2).

Acting as an ELR, the general government sector would thus provide jobs on an ongoing basis, to wit, immediately, to anybody who is not employed in the private sector economy. In particular, according to Wray (1998:127), the government might set up a “basic public sector employment” (BPSE) programme, in order for the unemployed to “be placed into a full time BPSE job to obtain on-the-job training”. Alternatively, it may enrol them in a part- or full-time educational programme. Both are viewed as increasing the employment level of the economy, the ultimate goals being to pay individuals (who otherwise would remain, at least for some time, unemployed) to work and produce, to promote efficiency of the private sector, and to increase the education and skill of ELR workers (see Harvey 1989).

Now, despite its appeal to theoreticians as well as domestic policy makers, a number of critiques can be, and have been, raised against the ELR view. In fact,

although Wray (1998) claims that job productivity, and the nature of produced output, are not ignored by a BPSE programme, the jobs provided under it aim at disposing of unemployment and not so much at providing social benefits in terms of output produced (Sawyer 2003:884). Further, “the ELR ‘solution’ appears to be to provide employment which does not require any capacity or skill” (p. 884). This solution, in fact, does not seem to consider (sufficiently) the problem of structural unemployment, that is, the mismatch between the skills looked for by firms and the skills offered by unemployed on the labour market (Kadmos & O’Hara 2000).

To be true, Wray (1998:142–3) lists some jobs that ELR workers could do, acting for instance as companions to the elderly, public school classroom assistants, safety monitors, neighbourhood cleanup engineers, day care assistants, library assistants, environmental safety monitors, artists, musicians, and so on. As Kadmos and O’Hara (2000:14) point out, however, the usefulness of these jobs in the private sector might be limited: “An ELR worker may well be gaining experience in a job that can never translate into private sector activity.” To allow ELR workers to take up a position in the private sector, hence to address structural unemployment problems, any ELR programme would have to make sure that the ELR jobs offered to the unemployed give them the skill and experience that will enable them to move to the private sector economy during the recovery phase of a business cycle (p. 13).

Now, even if ELR programmes were to be complementary to the private sector economy and provide to workers the skills required by the latter sector (say, in terms of information and communication technologies), a problem remains, because

these programmes critically depend on the TDM approach. Kadmos and O’Hara (2000:10) cogently note that this is the crucial point of any ELR policies, since “[t]he government is in a position to hire all unemployed workers at any price it chooses, financing this labour force by printing as much money as required that will achieve full employment” (see also Sawyer 2003:885–7). This point needs careful consideration, since it raises a number of crucial issues, which we consider in the remainder of this article.

### **Value of State Money**

A key assumption in the TDM approach is the state’s ability to issue debt (IOUs, fiat money) that has a final settlement power *per se*. “This means the [US] government can buy anything that is for sale for dollars merely by issuing dollars” (Wray 1998:ix). In fact, the purchase of private sector goods or services calls for a final payment. Payment finality means that “a seller of a good, or service, or another asset, receives something of equal value from the purchaser, which leaves the seller with no further claim on the buyer” (Goodhart 1989:26; see also Kahn & Roberds 2002). This is indeed problematic in the TDM case: the state obtains real goods and services, including the labour services provided by ELR workers, as a physical counterpart of nominal tokens (that is, bank notes and/or coins), which the state “fabricates” at a trifling cost—as notably the seigniorage view argues.

As a matter of fact, when the state pays for its purchases on the factor and goods markets with fiat money as in the TDM approach, it is merely surrendering to its creditors a promise to pay in the form of perfectly liquid financial claims (to wit, bank notes and reserves). This payment by

the state cannot be considered as final. It is indeed a mere promise of payment finality: it is only when the public disposes of fiat money at state pay offices, namely for the payment of tax liabilities, that the transactions between government and the private sector economy are cleared. Hence, payment finality occurs by some sort of barter trade: privately produced goods and services (including also labour services) are bartered against all sorts of tax obligations, with the intermediary of state (fiat) money. This is so because the seller of a good or of a (labour) service to the government actually pays its tax obligations only when this very same agent returns the corresponding amount of state (fiat) money to its issuer—and not before. In short, although money's validation can be explained in TDM terms, to wit, by taxation powers of the state, the state cannot determine money's value (that is, its purchasing power) by law. This value is established by the exchange of real goods, services, and financial assets in this context.

If an agent agrees to exchange part of his “initial endowments” of real goods, or labour services, for a number of money units issued by the state, he does so because he knows the value of the state money he receives and that he keeps as a temporary abode of purchasing power, since it will enable him to buy some other goods and services or pay his own tax obligations later. Now, how can he assess the precise amount of purchasing power of the received sum of state money, if the value of the latter depends on the terms with which his commodity is finally exchanged against some others or against some sort of tax obligations? In other words, what is the value of a (dollar-denominated) unit of state money?

According to Wray (2003a:91), “the ‘real’ value of the dollar will be determined by the ‘effort’ involved in obtaining it, that is, the labour services or basket of commodities one must provide to obtain a fiat money dollar.” This is a clear indication that a labour theory of value is required to determine, as well as to measure, this effort in terms of money. As Rossi (1999:484) notes, however, the TDM approach has yet to provide such a theory, which must associate money and production consistently with the bookkeeping nature of money (see Ingham 1996, 2000). Indeed, Wray (2003a:91) might be on the right track when he considers the state's role and goals. Although it requires to be developed, Wray's analysis of money's value provides, in fact, some clues in order to explain the purchasing power of what he calls state money.

Let us consider Wray's example (p. 91), to “presume that the state only wants to purchase labour services [...] and offers to pay a dollar of state notes per hour of labour services hired. Setting to the side obvious labour heterogeneity complications, the fiat money dollar will be worth an hour of labour.” Clearly, in this example, a one-dollar note will have the power to buy the result of an hour of labour, that is, the corresponding output. In light of Keynes's (1936:ch.4) concept of wage units, the dollar paid out by the state to its (ELR) workers for an hour of labour is indeed the monetary measure of the output produced by them over the same period of time. Generalising this principle, we may argue that each newly produced good or service is measured, in economic terms, by the number of money units paid out to those wage earners who produced it. This is so both in the public and private

sector economy (see Rossi 2001:122–31 for elaboration).

Now, assuming that the TDM approach is or can be grounded on the wage-units concept proposed by Keynes (1936), a major problem remains in this approach, since the TDM view assimilates state money with central bank money. In fact, these monies are distinct in the real world.

### **Distinction between State Money and Central Bank Money**

In the view of TDM advocates, “government keeps two sets of books, the Treasury’s book and the central bank’s book” (Wray 1998:79; see also Bell 2000). The central bank is then considered as the treasury’s bank, so much so that central bank money is state money, in the sense that the former money is issued by the central bank on behalf of the treasury. “For our purposes, it is not important to distinguish between the Fed’s and the Treasury’s balance sheet. The bank reserves carried on books as the bank’s asset and as the Fed’s liability are nothing less than a claim on government fiat money – at any time, the bank can convert these to coins or paper notes, or use them in payments to the state” (Wray 1998:78–9).

Now, as Gnos and Rochon (2002) argue, it is necessary to distinguish the function of the central bank from the role of the treasury. In fact, if the central bank acts as the treasury’s bank, this is only because the central bank can convert state debt (fiat money issued by the treasury according to the TDM approach) into bank money, and vice versa, every time a transaction occurs between the state and the private sector economy. Hence, it is wrong to do as TDM authors suggest, and “consolidate the central bank and the treasury, calling the conglomerate ‘the

state’, and combine treasury and central bank liabilities into a ‘high-powered money’ or ‘fiat money’” (Wray 2003a:91).

In fact, central bank money is not a state’s debt. Rather, it is the means of final payment at both the interbank and government levels. In nearly all countries, the government banks at the central bank, a privilege that a few big private sector firms are also endowed with in some nations. As Gnos and Rochon (2002:48) point out, “[t]he fact that state expenses and revenues each affect the amount of central bank money in the hands of commercial banks is not a sufficient reason to identify them [that is, state money and central bank money] with each other.” Further, since the central bank acts as final settlement agent—clearing debts in the domestic economy (for both the private and the public sectors)—state debt needs to be converted into central bank money in order for a final payment to take place between the state and the private sector. This is so even in the case when the treasury sells government bonds to the central bank (see Lavoie 2003:527, Table 21.15).

It should also be noted that in some cases the treasury does not actually issue money. In the United States, for example, the Treasury “pays its creditors from its account with the central bank, and thus uses central bank money” (Gnos and Rochon 2002:48). It is thus wrong to assimilate central bank liabilities to treasury liabilities, as TDM advocates do, and to “treat both as essentially ‘high-powered money’ or liabilities of the state” (Wray 2003a:87). This raises another critical note: if central bank money is different from the state’s liability, then, contrary to the TDM view, the public does not need to obtain state money in order for it to pay taxes. As a matter of fact, “taxes



are almost exclusively paid using bank money” (Wray 2003a:91), which shows that the latter money is important and needs to be carefully considered in any monetary analysis of modern production economies, with a state that imposes tax liabilities to the population.

### **Importance of Bank Money**

The TDM approach has been neglecting, so far, the importance of bank money in modern economies. In this approach, banks are just an afterthought. Indeed, they are not essential for the state theory of money. According to the TDM approach, fiscal policy is the source of money in an economic system of production and exchange: “money drops vertically to the private sector from government through government purchases of goods and services” (Wray 1998:111).

In fact, government cannot purchase goods from private sector agents before these goods have been produced, which requires firms to ask banks for credit in order to pay out wages to workers. This is a “first principle” valid for the private and the public sector as well. It is now well embedded in the monetary theory of production advocated by several endogenous money proponents (Graziani 2003). Clearly, even when the state needs to pay out wages to public sector workers, banks must grant to it a credit line on which the state can draw when the wage bill is paid. The fact that the state banks at the central bank does not change this analysis, because the central bank acts in this situation as any other bank does: it issues the means of payment for the state to finally pay its workers for the current production period. As a result, production precedes government spending logically as well as in point of fact.

Further, bank notes are not the prerogative of the state. As monetary history shows, there were many instances where private banks also issued notes (fiat money), which circulated in a more or less wide geographical area for the settlement purposes of the local community. For example, in the Middle Ages, particularly at the beginning of the fourteenth century, a number of periodical fairs were being held in Europe, the most famous and largest of them in Champagne, France, to which came traders and bankers from all over Europe. “Exchange booths were established and debts and credits were cleared to enormous amounts without the use of a single coin” (Innes 1913:396). To be sure, all private debt obligations were settled using private bank (goldsmiths’) notes, before the advent and expansion of clearing houses, which made the settlement process a mere bookkeeping affair. Also, in the case of the United Kingdom, “it took several banking acts – the 1826, 1833, and 1844 acts being the crucial ones – to limit the ability of banks to issue their own notes” (Rochon and Vernengo 2003:61–2). In addition, up to 1946 the Bank of England was a private institution, although with important ties with the Treasury. Even if today the central bank is part of the general government sector in all countries around the world, this does not mean that the latter agent can finance its expenditures by merely relying on the so-called printing press. As a matter of fact, central banks have been made more and more independent of governments, which has also resulted in the adoption of no-bail-out clauses as a general rule (like in the Treaty on the European Union).

The power of the state to tax and to define the unit of account is therefore not necessary for an economy to be monetised.

As Rochon and Vernengo (2003:61) argue, “firms will produce even if states are relatively weak, and hence unable to tax or force payment in a particular token.” To be sure, firms need bank credit to start a production process and pay out wages. In this respect, “[b]anks may issue bank notes, which are used by firms and accepted by workers, since workers know that they could use them to buy goods from the firms. Firms will accept bank notes since they have to repay their loans to the banks” (p. 61).

Money is thus “a creature of banks rather than a creature of the state” (p. 61). This was so even before the advent of “banks” as such: in ancient societies, in fact, goldsmiths acted as bankers act today, since they kept books in which they recorded all debts and credits for further reference and settlement. In this sense, “money is memory” (Kocherlakota 1998). As Innes (1913:407) pointed out, “[a] bank note differs in no essential way from an entry in the deposit register of a bank. Just like such an entry, it is an acknowledgment of the banker’s indebtedness [...]. The only difference between a deposit entry and a bank note is that the one is written in a book and the other is on a loose leaf”.

Courbis et al. (1991:329) clearly illustrate this point referring to the monetary history of the United Kingdom, notably, at the time of the first goldsmiths in London, around 1660–65. As they point out (pp. 324–5), book-entry payments existed long before bank notes, or their ancestors, say, a goldsmith’s certificates, appeared on earth. Further, like bank money, even fiat money is a form of credit-money. As a matter of fact, the economic foundation of any form of money is credit, not state (pp. 329–31). Fiat and bank monies pertain thus to the same category

(see Mehrling 2000), albeit fiat money, notably in the form of bank notes, increased and extended monetary circulation very much beyond those agents having a bank account. This, in fact, increased the size of money-wage economies, in which money and production are the two faces of the same coin, and in which banking and production systems define together the same structure of domestic and international exchange.

## **Conclusion**

The TDM approach is an alternative view on money with respect to traditional approaches. It argues that taxes are not used to finance the general government outlays, but are meant to sustain the demand for state money. With this view, it becomes possible, and even desirable, to implement public policies aimed at solving both inflation and unemployment problems. It is by following the Lerner (1943) principle of functional finance that the government will be able to obtain price level stability as well as full employment. If the state acts namely as an “employer of last resort,” it may provide a job to all those unemployed who are able and willing to work. This employment policy can be financed by running the “printing press” as the state can drain excess money, hence avoid inflation, by imposing tax liabilities to money holders (see Lerner 1947).

Clearly, this approach departs from more orthodox thought and policy. However, more work needs to be done to establish it as a true alternative to contemporary economic policy. In fact, modern central banks are independent of governments (particularly the treasury) and central bank money cannot be assimilated to a state debt. In addition, despite the fact that money is of a book-entry nature, the

idea that the government can provide full employment and price level stability at will is rather naïve (Rochon and Vernengo 2003:65). In fact, tax liabilities are not a means to avoid inflation. They give the state the power to issue bonds, hence to finance government spending (Gnos & Rochon 2002).

Properly understood, fiscal policy is thus a means to stabilise the economic system by the appropriate balance between taxes and public outlays in any phase of the business cycle. In this respect, monetary policy contributes to it, through its influence on the short-term rate of interest in the interbank money market, and in general on the financial market as a whole, to provide for a context of financial stability and economic prosperity in terms of real GDP as well as employment growth.

### Selected References

- Ahmed, Habib and Stephen M. Miller. (2000) "Crowding-Out and Crowding-In Effects of the Components of Government Expenditure", *Contemporary Economic Policy*, Volume 18, Number 1, pp. 124–33.
- Aschauer, David A. (1989) "Does Public Capital Crowd Out Private Capital?", *Journal of Monetary Economics*, Volume 24, Number 2, pp. 171–88.
- Barro, Robert J. (1990) "Government Spending in a Simple Model of Endogenous Growth", *Journal of Political Economy*, Volume 98, Number 5, pp. 103–26.
- Bell, Stephanie A. (2000) "Do Taxes and Bonds Finance Government Spending?", *Journal of Economic Issues*, Volume 34, Number 3, pp. 603–20.
- Bell, Stephanie A. (2001) "The Role of the State and the Hierarchy of Money", *Cambridge Journal of Economics*, Volume 25, Number 2, pp. 149–63.
- Bell, Stephanie A. (2003) "Neglected Costs of Monetary Union: The Loss of Sovereignty in the Sphere of Public Policy", in Stephanie A. Bell and Edward J. Nell (Editors), *The State, the Market and the Euro: Chartalism versus Metallism in the Theory of Money*. Cheltenham and Northampton: Edward Elgar, pp. 160–83.
- Bell, Stephanie A. and Edward J. Nell. (2003) (Editors), *The State, the Market and the Euro: Chartalism versus Metallism in the Theory of Money*. Cheltenham and Northampton: Edward Elgar.
- Copeland, Morris A. (1981) "Bank Deposit Currency Before A.D. 1700", *Research in Economic History*, Volume 6, pp. 245–54.
- Courbis, Bernard, Eric Froment and Jean-Michel Servet. (1991) "Enrichir l'économie politique de la monnaie par l'histoire", *Revue Économique*, Volume 42, Number 2, pp. 315–38.
- Easterly, William and Sergio Rebelo. (1993) "Fiscal Policy and Economic Growth: An Empirical Investigation", *Journal of Monetary Economics*, Volume 32, Number 3, pp. 417–58.
- Fernald, John G. (1999) "Roads to Prosperity? Assessing the Link between Public Capital and Productivity", *American Economic Review*, Volume 89, Number 3, pp. 619–38.
- Forstater, Mathew. (1998) "Flexible Full Employment: Structural Implications of Discretionary Public Sector Employment", *Journal of Economic Issues*, Volume 32, Number 2, pp. 557–63.
- Gnos, Claude and Louis-Philippe Rochon. (2002) "Money Creation and the State: A

- Critical Assessment of Chartalism”, *International Journal of Political Economy*, Volume 32, Number 3, pp. 41–57.
- Goodhart, Charles A.E. (1989) *Money, Information and Uncertainty*. Basingstoke and London: Macmillan, second edition.
- Goodhart, Charles A.E. (1998) “The Two Concepts of Money: Implications for the Analysis of Optimal Currency Areas”, *European Journal of Political Economy*, Volume 14, Number 3, pp. 407–32.
- Gordon, Wendell. (1997) “Job Assurance—The Job Guarantee Revisited”, *Journal of Economic Issues*, Volume 31, Number 3, pp. 826–34.
- Graziani, Augusto. (2003) *The Monetary Theory of Production*. Cambridge: Cambridge University Press.
- Harvey, Philip. (1989) *Securing the Right to Employment: Social Welfare Policy and the Unemployed in the United States*. Princeton: Princeton University Press.
- Ingham, Geoffrey. (1996) “Money is a Social Relation”, *Review of Social Economy*, Volume 54, Number 4, pp. 507–29.
- Ingham, Geoffrey. (2000) “‘Babylonian Madness’: On the Historical and Sociological Origins of Money”, in John Smithin (Editor), *What is Money?* London and New York: Routledge, pp. 16–41.
- Innes, A. Mitchell. (1913) “What is Money?”, *Banking Law Journal*, May, pp. 377–408.
- Kadmos, George A. and Phillip A. O’Hara. (2000) “The Taxes-Drive-Money and Employer of Last Resort Approach to Government Policy”, *Journal of Economic and Social Policy*, Volume 5, Number 1, pp. 1–22.
- Kahn, Charles M. and William Roberds. (2002) “The Economics of Payment Finality”, *Federal Reserve Bank of Atlanta Economic Review*, Volume 87, Number 2, pp. 1–12.
- Keynes, John M. (1930) *A Treatise on Money*, in *The Collected Writings of John Maynard Keynes*. London and Basingstoke: Macmillan, Volumes V–VI, 1971.
- Keynes, John M. (1936) *The General Theory of Employment, Interest and Money*, in *The Collected Writings of John Maynard Keynes*. London and Basingstoke: Macmillan, vol. VII, 1973.
- Knapp, Georg F. (1905) *The State Theory of Money*. Abridged edition translated by H.M. Lucas and J. Bonar. London: Macmillan, 1924.
- Kocherlakota, Narayana R. (1998) “Money Is Memory”, *Journal of Economic Theory*, Volume 81, Number 2, pp. 232–51.
- Lavoie, Marc. (2003) “A Primer on Endogenous Credit-Money”, in Louis-Philippe Rochon and Sergio Rossi (Editors), *Modern Theories of Money: The Nature and Role of Money in Capitalist Economies*. Cheltenham and Northampton: Edward Elgar, pp. 506–43.
- Lerner, Abba P. (1943) “Functional Finance and the Federal Debt”, *Social Research*, Volume 10, Number 1, pp. 38–51.
- Lerner, Abba P. (1947) “Money as a Creature of the State”, *American Economic Review*, Volume 37, Number 2, pp. 312–17.
- Mehrling, Perry. (2000) “Modern Money: Fiat or Credit?”, *Journal of Post Keynesian Economics*, Volume 22, Number 3, pp. 397–406.
- Miller, Stephen M. and Frank S. Russek. (1997) “Fiscal Structures and Economic

- Growth”, *Economic Inquiry*, Volume 35, Number 3, pp. 603–13.
- Mitchell, William F. (1998) “The Buffer Stock Employment Model and the NAIRU: The Path to Full Employment”, *Journal of Economic Issues*, Volume 32, Number 2, pp. 547–55.
- Mitchell, William F. and Martin J. Watts. (1997) “The Path to Full Employment”, *Australian Economic Review*, Volume 31, Number 4, pp. 436–44.
- Rochon, Louis-Philippe and Matias Vernengo. (2003) “State Money and the Real World: or Chartalism and Its Discontents”, *Journal of Post Keynesian Economics*, Volume 26, Number 1, pp. 57–67.
- Rossi, Sergio. (1999) “Book Review of Wray’s *Understanding Modern Money*”, *Kyklos*, Volume 52, Number 3, pp. 483–5.
- Rossi, Sergio. (2001) *Money and Inflation: A New Macroeconomic Analysis*. Cheltenham and Northampton: Edward Elgar.
- Rostovtzeff, Michael I. (1941) *The Social and Economic History of the Hellenistic World*. Oxford and New York: Clarendon and Oxford University Press.
- Sawyer, Malcolm. (2003) “Employer of Last Resort: Could It Deliver Full Employment and Price Stability?”, *Journal of Economic Issues*, Volume 37, Number 4, pp. 881–907.
- Smith, Adam. (1776/1976) *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford: Clarendon Press.
- Wray, L. Randall. (1998) *Understanding Modern Money: The Key to Full Employment and Price Stability*. Cheltenham and Northampton: Edward Elgar.
- Wray, L. Randall. (2003a) “Seigniorage or Sovereignty?”, in Louis-Philippe Rochon and Sergio Rossi (Editors), *Modern Theories of Money: The Nature and Role of Money in Capitalist Economies*. Cheltenham and Northampton: Edward Elgar, pp. 84–102.
- Wray, L. Randall. (2003b) “The Neo-Chartalist Approach to Money”, in Stephanie A. Bell and Edward J. Nell (Editors), *The State, the Market and the Euro: Chartalism versus Metallism in the Theory of Money*, Cheltenham and Northampton: Edward Elgar, pp. 89–110.
- Wray, L. Randall. (2004) “Conclusion: The Credit Money and State Money Approaches”, in L. Randall Wray (Editor), *Credit and State Theories of Money: The Contributions of A. Mitchell Innes*. Cheltenham and Northampton: Edward Elgar, pp. 223–62.

Sergio Rossi

Department of Political Economy  
University of Fribourg  
Switzerland  
sergio.rossi@unifr.ch

## Unemployment Compensation

*Oren M. Levin-Waldman*

### **Introduction**

In most industrialized nations, workers who lose their jobs through no fault of their own are generally entitled to receive unemployment compensation as means of wage replacement. Most unemployment compensation systems operate in the form of insurance systems. Workers qualify for insurance because they have worked a requisite period of time. Their employers have paid into a centralized Unemployment Insurance (UI) fund a certain percentage of their overall payroll. Because it is generally assumed that the premiums that employers pay are part of their workers' overall compensation—that the insurance is in lieu of higher wages—it is a benefit that workers have earned, and therefore is not contingent on need.

Nations in Europe began to experiment with unemployment compensation systems during the early part of the 20<sup>th</sup> century. France was the first country to establish a national level unemployment policy in 1905, and the United Kingdom was the first country to actually legislate a compulsory UI system with passage of the National Insurance Act in 1911. In the United States, UI grew out of the Great Depression during the 1930s when millions of workers found themselves out of work. But it was more specifically a response to the business cycle inherent to industrial production. Downturns in the business cycle would naturally result in layoff. And yet, if it was cyclical, workers could expect to be recalled back to their former employers once the business cycle picked up again. Therefore, unless there was a system in place to provide wage

replacement for laid off workers during the down cycles, there might not have been an available pool of workers to return upon recall.

UI has historically served two principal functions. First, it has offered laid-off workers critical income protection during temporary spells of joblessness. And second, to the extent it replaces lost wages it has fulfilled a macroeconomic of enabling workers to continue having purchasing power in the economy, especially during times of recession (Burtless 1991). Its crucial role is that it limits declines in consumer spending (Orszag 2001). As such it serves as an economic stabilizer, because by compensating for lost income, it has the macroeconomic effect of maintaining spending, and can thus offset the loss of national income (Rejda 1966). UI, then, seeks to alleviate hardship, prevent unemployment and promote reemployment by affording workers the time to search for new jobs that best match their skills with the demands of employers. And because it is an insurance program that workers have earned, claimants are able to maintain their self-respect (Blaustein et al 1993).

This essay will explore the concept of UI in terms of its history both in the United States and other industrialized nations, its theoretical construct, financing, and its effects. On one level, the UI system, was designed for an industrial economy subject to business cycles. In this system, layoffs were assumed to be temporary and in most cases workers were recalled to their employers. It was not necessarily designed to assist them with transitions from one type of labor to another. On another level, however, in a dynamic economy where jobs continuously undergo a process of what Joseph Schumpeter (1975) referred to

as “creative destruction,” layoffs are no longer temporary and workers need to make adjustments, which may entail a considerable overhaul of skills. Therefore, a system initially designed to ensure short-term income maintenance may simply be ill-equipped to meet the needs of a dislocated workforce which needs assistance in making transitions. This, of course, begs the question: is a system that was the outgrowth of the disruption caused by industrial production able to meet the needs of economies that continue to be in transition—from industrial/manufacturing based to post-industrial/information based? And yet, despite the changing nature of the economy, from which UI is supposed to provide some protection, one constant continues to remain: when workers lose their jobs, whether it be a temporary layoff or a permanent dislocation, they still need wage replacement. This is necessary not only for them and their families, but also their larger communities, because UI effectively minimizes the spread and depth of recession to the extent that it is able to stabilize the economy.

### **UI Systems**

All industrializing societies during the early part of the 20<sup>th</sup> century introduced unemployment compensation policies in response to the common experience of emerging industrial unemployment, but policies varied significantly in their industrial design. Therefore, a distinction must be drawn between UC which seeks to compensate workers for their lost wages through general public assistance programs and UI which specifically attempts to compensate them through a system of insurance. UI, then, is a particular form of unemployment compensation.

Policy makers were able to choose from a couple of different approaches: One was a means-tested policy of unemployment assistance in the tradition of existing poor relief programs. Another was to subsidize preexisting voluntary networks of unemployment compensation that had primarily been organized by trade union associations. This approach came to be known as the Ghent system and figured prominently in France’s construction of its national level policy when the French Parliament in 1905 established the Ghent system of UI “by introducing a government subsidy to all associations that offered benefits in case of involuntary unemployment (Mares 2000:223).” A yet third approach was to establish a contributory UI framework whereby financial and administrative responsibilities would be shared among unions, employers, and the state. But it was the Ghent system that would form the foundations for the type of compulsory UI system that would ultimately develop in much of the industrialized world.

What, then, does UI do and how does it work? Generally speaking, UI provides an initial line of economic defense for those who become unemployed through no fault of their own, and it spares them the indignation of public assistance. In broader terms, however, the concept of UI goes to the heart of just what role unemployment is to play in a vibrant economy. In other words, all free market economies require a certain degree of unemployment, and the question of UI only speaks to that degree. Gunther Schmid (1995) suggests that unemployment insurance is a social construct designed to solve the problem of uncertainty in the employment relationship—in short, to immunize workers from risk. Because unemployment

is part and parcel of employment, UI effectively becomes integral to the relationship. The nature of that relationship, then, is contingent on just how unemployment is viewed in relation to employment. Schmid actually suggests two different models, which also speak to some of the fundamental differences between UI in the U.S. and UI in many European countries. The first model is the market model, represented by what Schmid calls “hawks,” and it maintains unemployment to be a useful disciplinary device that forces workers to work harder and to accept both wage flexibility and spatial mobility. Therefore UI benefits should be low and should only exist as a short-term buffer. Otherwise, those who lose their jobs may not have an incentive to seek speedy reemployment. The other model is the welfare state model, represented by what he refers to as “doves,” and it maintains that since unemployment is part and parcel of the overall employment relationship there in fact needs to be active labor market policy. Unemployment benefits should only be resorted to as a last resort, but when given should be done so generously and with minimal restrictions. UI, in other words, is a social right that workers are entitled to by virtue of their citizenship in their respective countries. UI in the U.S., however, tends to emphasize individual responsibility. One has a right to it only because one has paid into the system. In Europe, unemployment compensation is seen more as a social obligation, and one has a right to it because one is a member of the community (Blaustein et al 1997).

Following these two different models, then, unemployment compensation may be governed by two different principles: either the insured person can be guaranteed a wage-related income or that person can be

guaranteed a minimum income. The U.S. tends to be modeled on the former while most European nations tend to more closely approximate the latter. The British system is perhaps the one most strongly governed by welfare state principles, emphasizing a minimum income guarantee rather than an insured guarantee of wage-related income. One is entitled to the minimum because one is a citizen. According to Jon Kvist (1998), the basis of provision in the liberal regime is need, and in the corporatist it is merit. That is, in a corporatist regime, one is entitled because one has effectively paid into the system, whereas in the liberal regime one merely has to be in need, in which case UI is more akin to another form of public assistance.

The concept of UI, in large measure, can be viewed as a halfway house between the market model of unemployment compensation, emphasizing greater individual responsibility and the welfare state model emphasizing greater social obligation. UI was not originally designed to be simply a welfare program for the indigent, rather it was to be an earned right for workers who became unemployed. It was specifically designed to be completely separate from depression-era relief programs. Unlike traditional public assistance where eligibility is based on need, eligibility is based on both labor force attachment and prior earnings experience. Need is merely presumed due to the economic loss that results from unemployment. The principal reason for why UI, at least in the U.S., was not based on demonstrated need, as is the case with public assistance, was to preserve the dignity of workers who were experiencing involuntary job loss. Since the beginning of the UI program in the U.S. there has been a general acceptance of the idea that



workers' benefits should replace one-half of workers' lost wages (O'Leary & Rubin 1997). UI, then, replaces lost income for those who lost their jobs through no fault of their own, and at the same time it serves to maintain some purchasing power during economic downturns. Moreover, it reduces dispersal of skilled employees when they are simply on temporary layoff, and thus helps to prevent a general breakdown of the labor market.

### **Eligibility**

In order to qualify for UI one has to have worked a certain number of weeks in the quarter prior to layoff, one has to have lost his/her job involuntarily, and one has to be ready and able to work. In all American states, and even many European countries, claimants must demonstrate that they are able and available for suitable work. What constitutes suitable work is determined by a mix of factors including the claimant's health, experience, training, prior earnings, and proximity to the claimant's residence. When a claimant does refuse to accept suitable work, claimants are either disqualified for a fixed number of weeks, a variable number of weeks, or for the duration of their unemployment. Disqualification for the duration actually reflects the assumption that unemployment should be involuntary and that if one refuses to accept suitable work which could end the unemployment spell, the unemployment is no longer involuntary, but has indeed become voluntary (Anderson 1997).

When UI was first established in the U.S., federal law only mandated that employers in industry and commerce be subject to UI taxes, and only if they employed a minimum of eight employees during at least a 20 week period. In 1954

coverage was extended to all commercial or industrial employers with a minimum of four workers. Coverage was expanded again in 1970 when new amendments required employers to pay UI taxes if they employed at least one worker working at least 20 weeks of the year or at a payroll of at least \$1500 in any of the calendar quarters. Coverage was also expanded to include employees in nonprofit organizations employing at least four workers. Then in 1976, coverage was yet expanded again to include some agricultural workers. Today UI is nearly universal with more than 90 percent coverage of all civilian employment in the U.S. (Bassi & McMurrer 1997).

Although UI in the US is governed by federal law, each state maintains its own UI system and establishes its own respective benefits and criteria for eligibility. That is, there is no unified national system, but 50 independent systems that are tied together by national guidelines. In most states, claimants only get benefits for 26 weeks and receive approximately 50 percent of their pre-layoff wages up to a maximum amount specified by their respective states. UI tends to be paid for out of payroll taxes from the employer. Payroll taxes are based on a percentage of their payrolls, and they are paid into state UI trust funds. What entitles workers to benefits as insurance is that they worked for these employers for a specified minimum number of weeks.

### **Comparative UI Systems**

How, then, does UI in the U.S. compare to other industrial nations, particularly those in the G-7? All seven nations have need-related assistance programs outside their unemployment compensation (UC) systems. By the end of the 19<sup>th</sup> century, unemployment protection schemes were

organized in several countries through trade unions, mutual benefit societies and other worker associations. During the early part of the Twentieth Century, these nations actually began to legislate compulsory UI programs. During the 1960s and early 1970s, UC programs were overhauled in several countries in response to changes in objectives held for UC. More emphasis was placed on integrating income maintenance aspects of UC with wider human resources policy, as well as greater job training and related provisions. Although UI in North America tends to be more market model oriented as opposed to welfare state oriented, UI in Canada tends to be more liberal than in the U.S. During the 1970s, benefits to unemployed persons other than job losers were significantly liberalized under Canada's Unemployment Insurance Act of 1971. In the U.S. benefits are granted primarily to job losers—workers on temporary and permanent layoff.

Under some limited circumstances, workers who quit for “just cause” can qualify, but generally job leavers, new entrants, and reentrants into the labor force do not qualify for benefits. In the U.S. where the work ethic still very much permeates American political culture, the objective is to provide sufficient relief and wage replacement without resulting in moral hazard. Ultimately, the objective is to get workers back to work; not to make it easy for them to forsake work for greater leisure. Because the mandate of Canada's law has been to assist “all individuals experiencing temporary earnings interruptions,” this has resulted in UI being more widely available to unemployed persons other than job losers (Moorthy 1989-90).

All seven nations use payroll taxes to fund their benefits. Tax rates in Italy are differentiated by industry, job type and firm size. In the U.S., rates vary by state and firm within states. In Japan, higher rates are applied to seasonal and construction jobs. And Germany relies most heavily on employee taxes with Canada and France next. All seven countries provide broad coverage to wage and salary workers. In France, Germany and Japan, coverage is actually coordinated with national pension systems by excluding those workers over pensionable age. The U.S. reduces benefits for those receiving pensions. In Canada, Germany and Japan, certain part-time workers are excluded from coverage based on a weekly threshold for hours worked, and Italy excludes from basic UI eligibility managerial personnel. In the U.S. and U.K. there are minimum work requirements based primarily on earnings which are relatively low. Each G-7 nation has rules disqualifying claimants whose unemployment results from voluntary quitting, misconduct, refusal of a suitable job, involvement in a labor dispute, or failure to accept training. All G-7 programs, except for the U.K. relate benefits to past wages. Germany replaces 63 percent of after-tax wages. Other systems base benefits on gross pay. Japan provides generous wage replacement for those towards the bottom of the distribution—80 percent replacement rate at low wage levels, which is also the highest replacement rate among the G-7. At the same time, Japan's wage figure also excludes overtime and bonuses. All G-7 countries require UI claimants to register with employment offices where information is provided on available jobs (Storey & Neisner 1997).

In Norway, Finland and Germany, not all G8 unemployed workers with dependents are actually entitled to higher benefits. The German UI system stresses that previous earnings and employment are guiding principles—the maximum duration of UI benefits increases with length of previous attachment to the labor market. Unemployed workers in Germany are transferred to unemployment support after their period of cash benefits through insurance rights has expired. In Denmark, Norway, and the U.K., claimants are transferred to cash assistance after their period of UI benefits has expired. In France, unemployed workers not entitled to insurance benefits are entitled to receive unemployment support for a year, during which time the allowance is at a fixed amount—higher for those over age 55—and afterwards the claimant is transferred to social assistance, which is essentially needs-based public assistance.

Some countries do not have UI. Australia, for instance, offers its unemployed workers a single benefit based on eligibility criteria similar to other countries. It is important to note that the experience of G-7 countries and other industrialized nations represents the importance of UC as a response to the fall out from industrializing economies. Because they underwent industrialization a long time ago, they had the opportunity to develop their UI systems. Therefore, it is no surprise that in many currently industrializing countries, or those in the transformative process of becoming market-based economies, attention has been focused on the need to develop UC systems based on insurance rather than social provision.

Following the fall of communism and the move to transform their economies into

market-based ones, John Micklewright (1991) observed that the emergence of unemployment that accompanied these moves towards market economies forced policymakers to look towards UC as a natural complement to the reforms that were taking place. Economic systems, like the former Soviet Union, that had once been characterized as right-to-work (as in one is entitled to a job) were now being replaced by ones in which employment carried risk. Whereas UI systems typically depend on employment history and are for unlimited duration, unemployment assistance (UA) does not. A state organized UI scheme represents an important demonstration of ‘solidarity’ among a workforce facing labor market upheaval.

Unemployment compensation schemes, however, have not existed in Eastern Europe for most of the post-WWII period, principally because of the individual’s right to work in a planned economy. While the former Soviet Union was hesitating over major reforms, it was widely recognized that there had to be a substantial re-allocation of labor and open employment. And as part of structural adjustment programs begun in early 1990, Poland introduced an unemployment benefits system that would pay 70 percent of previous wages for three months, then reducing to 50 percent for another six months, after which it would drop to 40 percent without limits in duration. Trade unemployment compensation was introduced in January 1989 paying a maximum of twelve months of benefits to those employed for at least 18 months during the previous three years. The initial rate was to be 70 percent of previous wages, falling to 60 percent after six months. To the extent that what was established in Poland is at all representative

of the Eastern European experience, it represents an attempt to strike a middle ground between UA and UI.

African countries, on the other hand, have not been in a position to set up systems of safeguards against unemployment. In Algeria, for instance, there has been a lack of clear policies for unemployment and economic revitalization. 2/3 of those unemployed are first-time jobless and have no working experience. Social protection for workers in Algeria has constantly evolved in response to problems arising from structural changes in the economy. Up until the beginning of the 1990s unemployment policies reflected both the pursuit of “full employment and jobs for life” and its limitations. With the introduction of UI, however, came market-based assumptions of just would motivate workers to seek reemployment. Monthly benefit levels were set on the basis of a reference wage equal to half of workers’ average wages over the last 12 months plus the national minimum wage. The duration was set at 2 months for every year of contribution and was to be no less than 12 months and no more than 36 months. But to encourage people to seek work, benefits were to be paid on a degressive basis in four installments: 100 percent of the reference wage would be paid during the first quarter; 80 percent of the reference wage would be paid during the second quarter; 60 percent of the reference wage would be paid during the third quarter and 50 percent would be paid during the fourth quarter. This degressive aspect was designed specifically to apply pressure on beneficiaries to seek work (Baussaidi 2003).

Most Latin American countries do not have UI; rather they seek to ameliorate the

hardships of unemployment through direct employment programs. While workers enjoy some coverage under social security systems, it tends to be inadequate due to problems in design and administration of systems, as well as the nature of the labor market. Instead of adopting passive labor market policies like UI, many Latin American countries have opted for active labor market policies such as direct employment, which attempts to reemploy workers. Direct employment programs aren’t a new feature in the region; rather they were used during the early 1980s in several Latin American countries to address the unemployment that arose from the adjustment crises of that decade. For those countries that have gone this route, most of the resources come from national budgets. Moreover, the evidence suggests direct employment programs have a positive effect for reducing unemployment and poverty, and that perhaps they provide a measure of stability similar to traditional UI (Reinecke 2005).

### **Labor Effects**

To the extent that UI immunizes workers from risk, it affords them greater opportunity to search for jobs with an appropriate match between the requirements of available jobs and the skills, education, and training of the unemployed. Theoretical studies show that with more generous benefits there is greater incentive for the unemployed to remain unemployed. The disincentive effect of UI is based on the premise that UI tends to prolong unemployment spells because it lowers the cost of unemployment. Those workers receiving UI benefits, then, tend to consume more leisure and reduce the intensity, i.e. the cost of a job search. UI, in short, makes them

more selective (Decker 1997). The income/leisure combination offered by most UI programs might be preferred by some recipients to the combination offered by a return to comparable employment. Therefore, some recipients may not be diligent in their job search for work, or they may adopt criteria for accepting employment that are unreasonably high from society's point of view. The result might then be a tendency to remain on the unemployment rolls for longer periods of time. On the other hand, by providing workers with replacement wages dislocated workers are afforded the opportunity to search for those jobs that best match their qualifications and skills. In this vein, UI effectively maximizes societal efficiency. From a macroeconomic perspective, an effective system of income protection will stabilize purchasing power. And from a social perspective, it will reduce societal divisions, thereby staving off strife and providing some justice to those who have lost their jobs through no fault of their own (Schmid 1995).

Neoclassical job search theory holds that UI, especially during longer periods of unemployment, effectively subsidizes job search, thereby leading to longer periods of unemployment (Jurajda and Tannery 2003). Much of the literature, however, also maintains UI to be a source of moral hazard (Feldstein 1978; Kelly 1979; Solon 1984; Katz and Meyer 1990; Meyer 1990). Because UI subsidizes the costs associated with the job search process, it makes it easier for workers to quit their jobs and thus spend time finding an alternative. So by making workers ineligible for benefits, UI systems along the U.S. market model strive to discourage voluntary quits.

Affected workers will also tend to demand less compensation for losing their

jobs, and employers are likely to feel less guilt for depriving them of their livelihood. According to conventional economic theory, generous wage replacement benefits reduce the probability of someone moving from unemployment (Feldstein and Poterba 1984; Meyer 1990). The effect is even stronger the longer the period over which unemployed workers are entitled to claim benefits. Higher UI benefits may actually have a strong negative effect on the probability of leaving unemployment (Meyer 1990). Conversely, the effect becomes weaker as the end of the benefit period approaches or as the benefit rate declines.

UI, however, reduces the probability of the unemployed moving into inactivity. The entitlement effects of UI increases the probability of those who aren't in the labor force joining it. Unemployment compensation may have a negative effect on the labor market because it causes the unemployed to be less willing to accept some job offers, as well as it may induce those in employment to quit to become unemployed. In other words, workers who know that they can rely on UI as a source of wage replacement have no real reason to fear losing their jobs, and may even welcome it if it allows them some leisure. This argument follows a traditional neoclassical assumption that workers do not work because they want to; they only do so because they have to. If they did not have to, they would simply trade off work for greater leisure. Some even attribute the rise in unemployment in Western Europe since the 1970s, and its persistence in a number of countries, to more generous levels of benefit payments.

Martin Feldstein (1978) has argued that UI particularly provides substantial incentive for increased temporary-layoff

unemployment; of all unemployment spells, temporary layoffs account for 50 percent. It essentially increases the duration of any given unemployment spell, and it may even induce more very short-term spells of unemployment. This is because employers are more willing to lay workers off when they are confident that they will be recalled. Similarly, Lawrence Katz and Bruce Meyer (1990) found that employer recall policies were the primary determinants of unemployment spell duration for individuals with non-negligible recall prospects. Those who expect to be recalled spend less time searching for jobs and have a lower finding rate than other UI recipients. Also the probability of leaving unemployment and finding new jobs increases greatly around the time that UI lapses. On the other hand, Gary Solon (1984) has suggested that the reduced availability of UI benefits to job quitters might be expected to reduce the frequency of quitting because it increases the expected costs of leaving employment.

These arguments only tend to be buttressed by the experience of Europe where welfare state benefits are more generous and unemployment is higher. Looking at both the Netherlands and the U.K., Elena Stancanelli (1999), found that most available empirical studies show that the hazard of exiting from unemployment rises when the level of unemployment benefits fades. In the U.K., for instance, the level of unemployment insurance paid is not contingent on previous earnings, rather it is a flat rate. Benefits in the U.K. actually approximate social/public assistance payments. Consequently as one passes from UI to social assistance, that person does not suffer a reduction in benefits. On the basis of data from the Survey of the Living Standard of the Unemployed

(LSUS) from June 21 to August 20, 1983—drawn from those unemployed that registered at Great Britain's Unemployment Offices—Stancanelli found that non-UI claimants did have shorter spells of unemployment than claimants. However, this was also due to their higher withdrawal rate from the workforce; because they withdrew from the workforce they were no longer counted among the unemployed. But there wasn't any significant evidence that those who were predicted to exhaust their benefits were necessarily more likely to exit unemployment.

Narendranathan et al (1985) looked at data constructed from the full DHSS Cohort Study of Unemployed, which involves a stratified random sample of unemployed men who registered as unemployed during the Fall of 1978. They constructed a longitudinal data set containing precise information on actual benefit receipt—an inflow into registered unemployment of 2300 men. With the length of first spell of registered unemployment as the dependent variable, the variable for employment income was the weekly net income from all sources, and real income was the measure of income while unemployed. There were marked differences in the impact of unemployment income across both age groups and durations. The impact for teenagers was considerably greater than for prime-age men, and the effect for older men was non-existent. The relationship between unemployment income and duration was found to be elastic, with greater income resulting in longer unemployment spells. Nevertheless, the effect of unemployment varied with age, and benefits had no impact on conditioned probability of leaving unemployment for

long-term unemployed. Moreover, the conditioned probability of leaving unemployment showed no signs of decreasing with duration, thereby providing strong evidence that reservation wages fall with duration. But there was very little evidence to support the view that benefit effects are greater for those whose benefit replacement ratio is already high.

Adding to this Schmid (1995) maintains that the assumption that more generous UI provisions will lead to longer spells of unemployment is not supported by cross-sectional comparisons of countries. On the contrary, because UI subsidizes the costs of searching for employment, it may become easier for workers to quit their jobs and spend more time finding alternative and possibly more suitable work. Robert Solow (1990) points out that to the extent that unemployed workers have choices, employment and job-search choices do not appear to be governed simply—or even peripherally—by those trade-offs between income and leisure of labor. Rather there must be other considerations, and the experiments claiming that incentives could reduce unemployment spells must entail factors which aren't being accounted for.

On the contrary, social institutions define acceptable and unacceptable modes of behavior in the most unlikely places, like the labor market. A persistent problem for economic theory has been persistent unemployment. An important difference between the market for labor and the market for other goods is that the performance of workers is contingent on the price paid for his/her services. The textbook model has a difficult time accommodating persistent unemployment because unemployed workers are supposed to put downward pressures on wage rates by competing for existing jobs. One would

expect that when there are a lot of unemployed workers, employers might actively solicit competitive wage cutting on their part, but this rarely happens. Nevertheless, there is a tension that policymakers need to be mindful of.

John Gal (2005) notes that policymakers in Israel used the “unemployment-trap” issue to justify changes in the country's UI program. When UI was first introduced during the early 1970s, policymakers sought to establish a program that would provide adequate protection to workers, while also encouraging them to return to work. Initially wage replacement ranged from 50 to 80 percent of previous earnings with maximum benefits duration in the program set at 138 days and 175 days for those over age 45. Participants in vocational training programs were also eligible to receive benefits during their training without forfeiting their right to the entire benefit duration period. Yet as unemployment declined during the late-1980s, public debate over issues related to the unemployment-trap grew more intense.

Decision makers, particularly in the Treasury ministry and the Bank of Israel, argued beginning in the early 1980s for cuts in benefit levels, duration of benefits, qualification conditions, and duration of qualification period as a means of dealing with the effects of the unemployment trap. But as much as these arguments appeared to garner public support which only resulted in tightening the program, it wasn't entirely clear that the rhetoric wasn't also being used as pretext to effectively cut the budget.

### **UI and Reservation Wages**

Martin Feldstein and James Poterba (1984) couch this disincentive effect in terms of

workers' reservation wages—the minimum wage that they will accept. They argue that the principal imperfection in modern labor markets is the downward rigidity of existing nominal wages. Consequently, the decline in the marginal value of the product of an employee's labor is likely to cause temporary layoff, as opposed to downward wage adjustment. Employees who lose jobs are likely to find that the wages at their next jobs are lower than the wages at their previous ones. Through a comparison of reservation wages with wages of last jobs, Feldstein and Poterba attempt to show the distortions caused by UI on the assumption that the probability of finding an acceptable job is likely to decline as the reservation wage exceeds the previous wage. In most cases the reservation wage will at least be equal to the previous wage. From a sample of unemployed individuals, the cumulative percentage of reservation wage rates was less than or equal to 62 percent of their previous wages. The remaining 38 percent of the sample had a reservation wage equal to or greater than their previous earnings.

As they see it, were it not for the current system of UI, individuals' reservation wages would of course be lower. Although UI reduces the cost of unemployment to the individual, it can nonetheless raise the unemployment rate in several quite different ways. For the individual who is unemployed and looking for a job, the lower cost of unemployment implies a higher reservation wage, and therefore a longer period of unemployment. Among those who are unemployed, the low potential cost of unemployment induces temporary layoff in response to reductions in product demand—even in response to seasonal fluctuations in employees' marginal revenue product (Feldstein and Poterba 1984).

Felstein and Poterba, however, may also assume too much with regards to the motivation of the unemployed. Perhaps UI does have the effect of raising the reservation wages of those at the low end of the wage distribution, or of anybody who can expect a replacement benefit of at least 50 percent of previous wages, but what benefit could there be for high income earners to sit at leisure at a fraction of previous pay over intensely searching for replacement of previous wages? UI more closely approximates the previous earnings of those at the lower end of the scale. Because each state imposes a maximum benefit, the higher the previous wage was, the less attractive UI becomes. Moreover, those at the upper end may also enjoy lucrative severance packages that offer them quite a bit of breathing room.

Surveys do show that many employers do indeed offer severance packages to their laid-off workers and that severance does indeed provide substantial income for many who have been displaced from long-term jobs (Kodrzycki 1998). What is not known from this study is who the unemployed are, i.e. their demographics. More than UI, which they really cannot establish as the source of one's reservation wage, prior income and living standards associated with that income might be a better source for one's reservation wage, as well as education, tenure in the labor force, and other considerations.

An alternative explanation is that those closed out of their jobs may simply be in a state of denial. During the early days of their unemployment, they do not really believe that they aren't going to return to their jobs. But as time passes, the reality sinks in. According to Gary Burtless (1990), the current knowledge of the impact of UI on labor supply is simply too



fragile. “Neither theory nor available empirical evidence permits us to predict unequivocally the net effect of unemployment insurance on labor supply.” By providing insurance to workers UI offers something of value to people who become unemployed. By supplementing the incomes of workers who become unemployed, it can slow down the process of reemployment. But without better empirical evidence than is currently available, it is impossible to predict which of these two basic effects will predominate. UI may increase the amount of economically productive job search and it might also raise the average productivity of workers by improving the match between jobs and workers. In situations where there are two job vacancies and two unemployed workers, it can be economically productive to subsidize the workers so that they can sort themselves into the two jobs that maximize their joint output and earnings. This isn’t to say that there aren’t adverse effects, but that the available evidence is simply insufficient to make any grand sweeping categorical conclusions (pp. 70, 82). In the end, the point of a social insurance system like UI is to reduce risk and maintain earnings.

### **UI Financing**

In many countries, UI benefits are financed by flat-rate payroll taxes, or out of general tax revenues. There is no connection between an individual firm’s behavior and its tax liability. In the United States UI is financed through an experience-rated payroll tax. Experience ratings are premiums levied against employers based on their history of layoff or the layoff pattern within their industries. Firms in industries with a greater probability of laying off their workers or firms with

histories of laying off workers are going to pay a higher percentage of their overall payrolls in UI taxes. Funds are deposited into state UI trust fund accounts maintained at the U.S. Treasury. The system of financing UI benefits in the U.S. is controlled at both the federal and state levels of government. To pay regular UI benefits, all employers are subject to a federal UI tax equal to 6.2 percent of their federal taxable income. A federal tax credit of 5.4 percent is available for those firms that have met a series of federal guidelines. These guidelines require that states have in place their own tax system with some form of experience rating, charging a lower tax rate to firms who lay off fewer workers.

Another component of the UI financing system subject to federal regulation is the taxable wage base. The federal government sets a minimum base—about \$7000 in 1997—which must be met by all states. Employers, then, only pay UI taxes on the first \$7000 of each worker’s earnings. States, however, may set a taxable wage base above the federal minimum. A major deficiency in the current system of UI financing is that infrequent, ad hoc adjustments to the taxable wage base lead to the continual erosion of its financial stability. Moreover, in more recent years, there have been dramatic declines in UI trust funds, which means that they are more likely to experience a deficit during a major recession—periods when UI benefits are likely to be extended beyond the typical 26 weeks (Levine 1997). Nearly half of UI programs required loans during the 1975-78 period, and total borrowing exceeded \$5 billion. The key factors making UI benefit financing complex and important are that benefit outlays are highly sensitive to the business cycle and that the timing and severity of the cyclical downturns are

difficult to predict (Miller, Pavosevich and Vroman 1997). Firms more likely to lay their workers off are bound to pay higher taxes, but experience rating in the UI system is still not quite the same as the idea of experience rated premiums in the general insurance literature. An imperfect rating of taxes, then, is likely to encourage unemployment. Layoffs generate income for a firm's workers with no corresponding cost for employers, thereby creating an incentive to compensate workers with UI rather than earnings. Unemployment thus becomes relatively more attractive, because it implies that both the incidence and duration of temporary layoff spells is increased (Topol 1990; Anderson and Meyer 1993).

Most European countries, however, do not rely on this type of system of experience rating, rather UI is considered to be a larger social cost. In Sweden, the costs of unemployment are shared by the central government, the pension insurance system, and the local and provisional governments. In Austria, the unemployment insurance fund is responsible for all expenditures on active labor market policy, but bears only about half of the fiscal burden of the unemployed. The Federal Republic of Germany represents an intermediate position, with 2/3 of the costs of unemployment being borne by tightly interrelated budgets of the federal government and the Federal Employment Institute (Schmid 1995).

### **UI and Labor Market Policy**

How, then, should we view UI within the context of a changing economy? Aside from income support, what role should the system play? By design, UI generally provides little more than wage replacement

and may actually encourage layoffs. Aside from measures aimed at encouraging job search, and in some cases making it a productive search, UI really isn't conceived of as a broader labor market policy. If employers know that their workers can fall back on UI, it then becomes much easier for them to dismiss workers under a UI regime. Consider for example that in the typical system in the U.S. and Canada, unemployed workers only receive benefits if laid off, but nothing if hours are simply reduced. By contrast, the typical European UI system routinely pays short-term compensation (STC) to those workers on reduced hours. Instead of laying off workers, employers can simply reduce work weeks, with their employees drawing partial benefits. The effect of STC, then, is to encourage work sharing rather than layoff, which tends to be characteristic of the American system (Burdett and Wright 1989).

This raises the question: are there ways that UI could be fine-tuned so that it would fit into a broader labor market policy? In 1993, the U.S. Congress passed the "Unemployed Compensation Amendments" which require new UI claimants upon filing for benefits to be profiled according to their demographic characteristics and work history. The purpose was to identify those claimants most likely to exhaust their regular UI and to then target them for job search assistance. The legislation was actually based on demonstration projects conducted in a few states during the 1980s which found that dislocated workers receiving job search assistance did find work from one half week to four weeks more quickly. The average reduction in most states was about one week, and there also appeared to be no effect on weekly wages (U.S. Department

of Labor 1995). On one level, such reforms are viewed as early intervention on behalf of the unemployed. But on another level, they could be said to represent first steps towards the integration of traditional UI with broader labor market policy.

Worker profiling essentially works as follows: the first UI payment triggers the profile which is based on the following criteria: recall status, union hiring hall agreement, education, job tenure, change in employment in previous industries, change in employment in previous occupations, and local unemployment rate. Each of these factors speak to the probability of the unemployment spell being a long one. Claimants either on recall or covered by union agreement are usually excluded. Those remaining are then assigned a probability of long-term unemployment on the basis of the profile. Then to the extent that services are available, “identified” claimants are either immediately referred to service providers or placed in a selection pool from which a referral may later be made. Services begin with an orientation session advising claimants on the availability and benefit of reemployment assessment, and if appropriate an individual assessment of each claimant’s needs is made. Then based on an individual service plan, which is considered to be a compact between the claimant and service provider, the claimant is referred to reemployment services tailored to the individual’s needs. Services are generally in the form of workshops on writing resumes and interviewing, as well as identifying whatever other needs the displaced might have so that they can become more marketable. And those referred are also required to participate as a condition for further receipt of UI benefits (Department of Labor 1994).

## **Conclusion**

Unemployment compensation has been evolutionary and continues to be because the nature of the general employment relationship continues to evolve. Although UI systems do more than they used to in terms of general labor market policy, UI continues to be at root an insurance system designed to provide workers with temporary and partial wage replacement and ultimately continued purchasing power. Regardless of how much the economy changes, workers will always have this basic need. At issue is whether a system that only provides limited coverage, initially predicated on the assumption that workers would be recalled upon an improvement in the business cycle, is equipped to deal with transitional economies. Workers who aren’t to be recalled may then need to make career changes, which require the development of skills. To the extent that UI in the US now profiles unemployed workers, it has taken a first step in that direction.

For UI systems to be suitable to the needs of changing economies, they must do more than merely provide income maintenance on the outdated assumption that whatever jobs are lost will simply return with changes in the normal business cycle. UI systems have to recognize that unemployment is not the homogeneous category it has been assumed to be during its inception in the early part of the Twentieth Century. Rather employment in the global economy has assumed different forms with more flexible wage systems, hours and general working arrangements. An effective UI system is one that is similarly flexible in order to accommodate these changes. At the same time, a UI system is an essential ingredient for the

maintenance of stability in a market economy. And that it serves as an effective automatic stabilizer may for the moment be good enough for some countries.

### **Selected References**

- Advisory Council on Unemployment Compensation. (1995) *Unemployment Insurance in the United States: Benefits, Financing, Coverage*. Washington DC: A Report to the President and Congress.
- Anderson, Patricia M. (1997) "Continuing Eligibility: Current Labor Market Attachment." in Christopher J. O'Leary and Stephen A. Wandner (Editors), *Unemployment Insurance in the United States: Analysis of Policy Issues*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- Anderson, Patricia M. and Bruce D. Meyer. (1993) "Unemployment Insurance in the United States: Layoff Incentives and Cross Subsidies." *Journal of Labor Economics*, Volume 11, Number 1, pt 2, pp. S70-S95.
- Atkins, Anthony B. and John Micklewright. (1991) "Unemployment Compensation and Labor Market Transitions: A Critical Review." *Journal of Economic Literature*, Volume 29 (December), pp. 1679-1727.
- Bassi, Laurie J. and Daniel P. McMurrer. (1997) "Coverage and Reciprocity: Trends and Effects." in Christopher J. O'Leary and Stephen J. Wandner (Editors), *Unemployment Insurance in the United States: Analysis of Policy Issues*. Kalamazoo, MI: Upjohn Institute.
- Boussaidi, Ali. (2003) "Unemployment Insurance in Algeria: Help or Hindrance?" *International Social Security Review*, Volume 56, Numbers 2-4, pp. 157-175.
- Blaustein, Saul J., Christopher J. O'Leary and Stephen Wandner. (1997) "Policy Issues." in Christopher J. O'Leary and Stephen J. Wandner (Editors), *Unemployment Insurance in the United States: Analysis of Policy Issues*. Kalamazoo, MI: Upjohn Institute.
- Blaustein, Saul J. with Wilbur Cohen and William Haber. (1993) *Unemployment Insurance in the United States: The First Half Century*. Kalamazoo, MI: W.E. Upjohn Institute.
- Burdett, Kenneth and Randall Wright. (1989) "Unemployment Insurance and Short-Time Compensation: The Effects of Layoffs, Hours per Worker, and Wages." *Journal of Political Economy*, Volume 97, Number, pp.1479-1496.
- Burtless, Gary. (1991) "The Tattered Safety Net." *The Brookings Review*, September, pp. 38-41.
- Burtless, Gary. (1990) "Unemployment Insurance and Labor Supply: A Survey." in W. Lee Hansen and James F. Byers (Editors), *Unemployment Insurance: The Second Half-Century*. Madison: University of Wisconsin Press.
- Decker, Paul T. (1997) "Work Incentives and Disincentives." in Christopher J. O'Leary and Stephen J. Wandner (Editors), *Unemployment Insurance in the United States: Analysis of Policy Issues*. Kalamazoo, MI: Upjohn Institute.
- Feldstein, Martin and James Poterba. (1984) "Unemployment Insurance and Reservation Wages." *Journal of Public Economics*, Volume 23, pp. 141-167.
- Feldstein, Martin. (1978) "The Effect of Unemployment Insurance on Temporary Layoff Unemployment", *American Economic Review*, Volume 68, Number 5, December, pp. 834-846.
- Gal, John. (2005) "The Rise and Fall of Unemployment Insurance in Israel." *International Social Security Review*, Volume 58, Number 4, pp.107-116.
- Gangl, Markus. (2004) "Welfare States and the Scar Effects of Unemployment: A Comparative Analysis of the United States and West Germany." *American Journal of*

- Sociology*, Volume 109, Number 6, May, pp. 1319-64.
- Jurajda, Stepan and Frederick J. Tannery. (2003) "Unemployment Durations and Extended Unemployment Benefits in Local Labor Markets", *Industrial and Labor Relations Review*, Volume 56, Number, January, pp. 324-348.
- Katz, Lawrence F. and Bruce D. Meyer. (1990) "Unemployment Insurance, Recall Expectations, and Unemployment Outcomes", *Quarterly Journal of Economics*, November, pp. 973-1001.
- Kelly, William J. (1979) "Theoretical Foundations for a Reform of Unemployment Insurance", *Social Science Quarterly*, Volume 60, pp. 96-104.
- Kodrzycki, Yolanda K. (1998) "The Effects of Employer-Provided Severance Benefits on Reemployment Outcomes." *New England Economic Review*, November/December, pp. 41-68.
- Kvist, Jon (1998) "Complexities in Assessing Unemployment Benefits and Policies", *International Social Security Review*, Volume 51, Number 4, pp. 33-55.
- Levine, Phillip B. (1997) "Financing Benefit Payments." in Christopher J. O'Leary and Stephen J. Wandner (Editors), *Unemployment Insurance in the United States: Analysis of Policy Issues*. Kalamazoo, MI: Upjohn Institute.
- Mares, Isabela. (2000) "Strategic Alliances and Social Policy Reform: Unemployment Insurance in Comparative Perspective", *Politics & Society*, Volume 28, Number 2, pp. 223-244.
- Meyer, Bruce D. (1990) "Unemployment Insurance and Unemployment Spells", *Econometrica*, Volume 58, Number 4, pp. 757-782.
- Micklewright, John. (1991) "The Reform of Unemployment Compensation: Choices for East and West", *European Economic Review*, Volume 35, pp. 419-429.
- Miller, Mike, Robert Pavosevich and Wayne Vroman. (1997) "Trends in Unemployment Benefit Financing", in Christopher J. O'Leary and Stephen J. Wandner (Editors), *Unemployment Insurance in the United States: Analysis of Policy Issues*. Kalamazoo, MI: Upjohn Institute.
- Moorthy, Vivek. (1989-90) "Unemployment in Canada and in the United States: The Role of Unemployment Insurance Benefits", *FRBNY Quarterly Review*, Winter, pp. 48-61.
- Narendranathan, W., S. Nickell and J. Stern. (1985) "Unemployment Benefits Revisited", *The Economic Journal*, Volume 95, June, pp. 307-329.
- O' Leary, Christopher J. and Murray A. Rubin. (1997) "Adequacy of the Weekly Benefit Amount." in Christopher J. O'Leary and Stephen J. Wandner (Editors), *Unemployment Insurance in the United States: Analysis of Policy Issues*. Kalamazoo, MI: Upjohn Institute.
- Orszag, Peter. (2001) *Unemployment Insurance as Economic Stimulus*. Washington DC: Center for Budget and Policy Priorities. November 15.
- Rejda, George E. (1966) "Unemployment Insurance as an Automatic Stabilizer", *The Journal of Risk and Insurance*, Volume 33, Number 2, pp. 195-207.
- Reinecke, Gerhard. (2005) "Income Protection Through Direct Employment Programmes: Recent Concepts and Examples from Latin America." *International Social Security Review*, Volume 58, Numbers 2-3, pp. 163-183.
- Schmid, Gunther. (1995) "Institutional Incentives to Prevent Unemployment: Unemployment Insurance and Active Labor Market Policy in a Comparative Perspective", *The Journal of Socio-Economics*, Volume 24, Number 1, pp. 51-103.

- Schumpeter, Joseph A. (1975) *Capitalism, Socialism and Democracy*. New York: Harper and Row.
- Solon, Gary. (1984) "The Effects of Unemployment Insurance Eligibility Rules on Job Quitting Behavior", *The Journal of Human Resources*, Volume 19, pp. 118-126.
- Solow, Robert M. (1990) *The Labor Market as a Social Institution*. Cambridge, MA: Basil Blackwell.
- Stancanelli, Elena G.F. (1999) "Unemployment Duration and the Duration of Entitlement to Unemployment Benefits: An Empirical Study for Great Britain", *Applied Economics*, Volume 31, pp. 1043-1051.
- Storey, James R. and Jennifer A. Neisner. (1997) "Unemployment Compensation in the Group of Seven Nations", Christopher J. O'Leary and Stephen J. Wandner (Editors), *Unemployment Insurance in the United States: Analysis of Policy Issues*. Kalamazoo, MI: Upjohn Institute.
- Topol, Robert. (1990) "Financing Unemployment Insurance: History, Incentives, and Reform." in W. Lee Hansen and James F. Byers (Editors), *Unemployment Insurance: The Second Half-Century*. Madison: The University of Wisconsin Press.
- U.S. Department of Labor. (1995) *What's Working (and What's Not)*. Washington DC: Government Printing Office.
- U.S. Department of Labor. (1994) *The Worker Profiling and Reemployment Service System: Legislation, Implementation Process and Research Findings*. Unemployment Insurance Occasional Paper 94-4. Employment and Training Administration.

Oren M. Levin-Waldman  
 Metropolitan College of New York  
 New York City  
 OLevin-Waldman@mcny.edu

## Workplace Agreements

*John Burgess*

### **Introduction**

All workplaces are subject to a set of arrangements, rules and understandings governing work. These may be formal or informal, codified or embedded in the collective memory. They may be enforceable or, a set of loose arrangements. These arrangements, rules and understandings can cover a multitude of issues including, starting and completion times, workplace safety arrangements and reward systems. The examination of workplace agreements highlights the importance of institutions and systems of formal and informal procedures for the organisation of work, the resolution of disputes, the determination of pay and benefits and mechanisms for implementing change at the workplace. It is the social and personal of work, of relationships at the workplace, of the rules and regulations governing the workplace, and formal and informal distribution mechanisms that are neglected in the neoclassical models of the labour market.

Employment engagements are ongoing and often long-term, with an implied and unwritten contract of employment, job tasks are often imprecise and changing, employment shifts can be flexible, and reward systems have to consider both individual and collective contributions. Also, the workplace has to operate in the context of legislative and community norms governing the timing and duration of work, minimum pay, recognition of parties to the agreement, procedures for dispute resolution and other standards such as occupational health and safety, non discriminatory employment arrangements

and environmental codes governing production at the workplace.

There is a trend towards decentralisation in industrial relations bargaining arrangements across the OECD and towards more fragmented bargaining arrangements (Visser 2001). The spread of neo-liberal ideas has been influential in shaping the nature of bargaining regimes and a number of Anglo Saxon economies in particular have shifted away from centralised and industry bargaining systems to decentralised bargaining systems based on the workplace. In the UK this has been accompanied by the shift towards the individualisation of employment contracts and attempts to exclude trade unions from the bargaining process (Brown and Ryan 2003). Although the trend towards decentralisation, and hence the workplace focus, has also been strong right across the EU (Locke & Kochan 1998; Visser 2001). One significant area of change has been the public sector where practices such as privatisation and outsourcing have had a considerable impact on the terms and conditions of employment and on the processes of industrial relations (Fairbrother et al 2002; Macdonald & Burgess 1999).

This article will consider the following: what is a workplace agreement, why have workplace agreements, what goes into workplace agreements, who determines the content of agreements, how are workplace agreements reached, should the workplace be the sole focus for agreement making, does trade liberalisation and the forces of globalisation lead to a converge in agreement making processes across countries and what is the broader impact of workplace agreement making on the economy? Finally, the article considers

research issues associated with workplace agreement making.

### **Workplace Plus Agreement**

The concept of the workplace is ambiguous. It is a location at which production and employment takes place within the one organisation. It does not necessarily equate with an organisation since it is possible to have multiple workplaces within the one organisation, the larger the organisation the greater the number of workplaces. It equates with the concept of an establishment, a sub-set of locations within an organisation. A workplace can be vast, as in the case of the public service, or it can be a single small business location. It may also entail a location where there are multiple employers such as a construction site. There is no uniformity as to what exactly is the workplace. To some extent it is the agreement that sets out to establish the working conditions and reward systems within the defined workplace. The location may be specific or it could apply to a number of different locations.

In the same context there is also ambiguity surrounding what is meant by an agreement. It suggests an arrangement that has been reached between managers and employees that govern most aspects of work, including reward systems and methods of dispute resolution. However, it is possible that much work takes place without a formal employment contract or a formal workplace agreement being present. The informal and the implicit are common in small business workplaces and for non-standard employment arrangements such as contact work and temporary employment where ambiguous employment status may exclude such workers from any formal bargaining process (Allen and Henry

1996). Indeed Allen and Henry (1996) suggest that there is a tendency in the UK towards a new employment regime characterised by individualised employment contracts and precarious employment arrangements, even in the traditionally secure and stable public sector. Not all workers performing duties at the workplace are necessarily covered by a workplace agreement. Exceptions include managers, contractors and agency workers. In addition, not all workplaces possess their own industrial agreement; their operations may be subsumed into an enterprise, industry or national agreement.

### **Parties to a Workplace Agreement**

An agreement is reached between parties, so who are the parties of a workplace agreement? First, there are the managers of the organisation. They may be the owners of the business or representatives of the owners of the business, in other cases, such as not for profit businesses, they may be representatives of a managing board, and finally, they may be appointed public sector managers. Second, there are the employees of the organisation who may be represented by a trade union, an elected representative, or by a delegated bargaining agent. Third parties may be respondents to the agreement, for example employer organisations or industrial tribunals, where agreements have to be ratified or certified.

For any workplace there may exist a multitude of agreements. It is not necessary that each workplace would have one agreement to cover all employees. There could be separate agreements for employees of different occupations or with different unions who represent the different occupations. Indeed, multiple union agreements are also possible. The agreement may be collective, bargained



between management and a group of employees or their representative, or they may be with individual employees. If there are individual workplace agreements then potentially there are as many agreements as there are employees.

It is debateable whether the formal parties to the agreements are those that determine the content and final outcomes of agreements. How autonomous workplace representatives are in the agreement process will be determined by a number of factors including the ownership structure of the workplace (stand alone organisation, branch office, subsidiary, public or private, degree of employee equity) and the structure and forms of employee representation (non union local representation, single or multiple unions, enterprise or craft unions, works councils). The issue of autonomy in agreement making is also questionable. For example, local managers may be subject to directives from head office, public sector organisations may be subject to government policies, and local trade union representatives may be subject to the guidelines laid down by regional or national union organisations. The early British workplace industrial relations studies revealed that bargaining was very centralised in the public sector, with little scope for local autonomy in decision making. For plant level negotiations in the private manufacturing sector consultations with head offices or higher management proceeded the majority of workplace negotiations by managers (Millward & Stevens 1986).

If the workplace is part of a larger organisation then local autonomy will be limited. In turn there are other constraining factors that limit workplace autonomy, these include the directives of employer

groups and trade union peak councils, the directives of government, and the directives of supra national organisations including international corporate head offices in the case of local subsidiaries and, in the case of the EU, the social charter. That is, systems of employment regulation, representation and bargaining will shape the nature and characteristics of bargaining at the workplace (Bamber & Lansbury 1987).

### **Why Have Workplace Agreements?**

Agreements set out conditions of employment, arrangements for settling disputes and grievances, who undertakes bargaining, job descriptions and responsibilities and remuneration arrangements. The agreements set out the rules and procedures governing employment at the workplace. An agreement usually represents a codified set of institutional rules and procedures that have been developed through time. The agreement is the formalisation of institutional rules governing work, the workplace, bargaining, employment conditions and dispute resolution. Having an agreement means that there are already procedures in place to deal with disputes, to set out how bargaining should proceed and establish what pay rates are etc. The agreement may not be legally enforceable and it may only set out custom and practice, but it does operate as a focal point for managers and employees in their day to day interactions at the workplace. Without an agreement then there is the potential for either extreme managerial prerogative to develop, or there is the prospect of an ad hoc process of bargaining and negotiation developing. In this context having an agreement can generate potential savings to

all parties in terms of reducing information and transaction costs.

The agreement can serve as a form of legitimisation of forms of managerial control, particular work practices or a given trade union organisation. That is, the agreement is not important for its content but for what particular forms of authority and control it legitimises (Armstrong et al 1981:ch.3). Behind the legitimisation can be often found an underlying set of ideological beliefs such as the rights of management to determine working conditions or the rights of a particular union to have priority over other unions in hiring. The agreement, and the assumptions behind the agreement, can be subject to forms of contestation. These can take the form of explicit industrial action and bans through to individual and informal forms of resistance such as work to rule, refusal of over-time, absenteeism and quitting (Deery & Plowman 1991:ch15).

### **What is in Workplace Agreements?**

An agreement will generally include the following components:

- The parties to the agreement
- Who is covered by the agreement and for what period
- The conditions of the agreement
- Processes for dispute resolution
- The terms of the agreement

The parties to the agreement are generally the owners of the workplace—rather than the management of the workplace. Employees may be covered by a union, or several unions, or forms of non-union representation. Each employee could be an individual signatory to a contract. There may be third parties to the agreement, including employer associations, trade union councils, industry

associations and industrial tribunals (Deery & Plowman 1991:Part 3).

The agreement will generally not cover all employees at a workplace. For example, managers may be subject to different terms and conditions of employment from other employees. Also, the agreement may only cover one section of a workplace or one trade union at a workplace. That is, the agreement may not be exhaustive or exclusive in terms of its coverage.

There is no pro forma of what should be in a workplace agreement, often the contents will be determined by precedent (what has been included in the past) and fulfilling required statutory and other legal obligations—for example, anti-discrimination, minimum pay rates etc. An agreement could be expected to cover the following:

- Qualifications, duties and responsibilities
- Training
- Terms and conditions of employment
- Termination of employment
- Wage rates
- Allowances
- Forms of consultation, including works councils
- Hours of work and organisation of work rosters
- Leave entitlements: sick leave, holiday leave, parenting leave etc
- Dispute and grievance resolution procedures
- Occupational health and safety arrangements
- Union recognition clauses or other forms of representation

The 1995 Australian Workplace Industrial Relations Survey reported that the three major issues negotiated with workplace delegates were work practices,

pay rates and working hours (Morehead Steele, Alexander *et al.* 1997:200). The 1998 British workplace survey reported that the main issues negotiated with trade union representatives were pay and conditions of employment, grievance resolution and systems of payment. With non trade union representatives the three main issues were occupational health and safety, equal employment opportunity programs and grievance resolution (Cully *et al* 1999:104).

These were part of a process of on going consultation and negotiation occurring at workplaces in Australia and Britain. That is, having a workplace agreement in place does not remove the need for ongoing workplace negotiation. This may be related to making the agreement operational, or, in many cases, it is associated with dealing with daily developments in the workplace such as occupational health and safety issues and addressing employee grievances.

### **Who Bargains and Who Determines the Content and Outcome of Agreements?**

The dimensions and characteristics of bargaining can vary across workplaces, industries and nations. Systems of law and regulation, the structure of employer organisations and trade unions, and forms of resistance and co-operation all interplay to determine the shape and nature of the bargaining process. Whether bargaining systems are converging in response to common pressures such as globalisation and technological change, or whether they are becoming more divergent and fragmented has been one of the enduring debates in industrial relations (Bamber & Lansbury 1987).

There may be a designated process in place that nominates the bargaining agents

and the process and procedures associated with this process. This may be formalised or it could be informal. The number of parties to the agreement may also be detailed or it could be a case of management and employees nominating representative bargaining agents. In some countries such as Australia there is a formal process of state provided conciliation and arbitration procedures that provide a framework for voluntary settlement and where this is absent, enforceable settlement (Deery & Plowman 1991:ch.4).

Does trade union presence have any impact on the bargaining outcome? There is an extensive literature on union and non-union wage differentials (Booth 1995; Freeman 1989; Macpherson & Stewart 1990) and on the impact of unions on other aspects of workplace performance including productivity and labour turnover (Freeman & Medoff 1984). In addition there are other issues surrounding trade unions and how they impact on the workplace, these include forms of unionism (industry, craft, enterprise), inter-union relations and politics, workplace activism and union politics (Booth 1995; Lansbury & Macdonald 1992:15).

The process of bargaining and the determination of bargaining outcomes has long been analysed and modeled by economists and mathematicians. Early pioneers in bargaining theory included von Neuman and Morgenstern (1944) and Nash (1950). The bargaining process could be modeled as a game between two or more players with set objectives, constraints and strategies. The games could be rival, co-operative, incorporate different strategies and time horizons, different pay-offs and costs. What game theory does is to formally capture the bargaining process,

such as that associated with wage contract bargaining, and set out how outcomes can shift according to shifts in such factors as strategies and coalitions associated with the bargaining process.

One characteristic of workplace bargaining as a game process is that it is often constrained by rules governing process and procedures, and constrained by industry norms shaped by the competitive environment or imposed by agreements reached between employer organisations and unions, or even more aggregative collective norms as established through national collective arrangements that include government – for example, centralist incomes policy arrangements (Visser, 2001). That is the role of the State as a determinant of the bargaining process and the bargaining outcomes is important in many countries.

### **Workplace Bargaining and Macroeconomic Performance**

An important policy issue surrounds the efficiency and effectiveness of workplace bargaining as opposed to more centralised bargaining at the industry or national level (Brown & Ryan 2003). The workplace may lead to more individualised and customised agreement making that better matches the economic circumstances of the workplace. However, there are much larger transaction and information costs associated with more frequent and diverse bargaining at the workplace when industry or national bargaining can reduce these costs and also institute industry or national norms with respect to outcomes.

Calmfors and Driffl (1988) suggested that highly decentralised bargaining arrangements and highly centralised bargaining arrangements were capable of delivering favourable macro-economic

outcomes relative to the intermediate or hybrid case (neither highly centralised or decentralised). In a simple plot of the unemployment outcomes of countries cross ranked along a ‘bargaining structure’ continuum (with highly decentralised at one end and highly centralised at the other), Calmfors and Driffl found a hump-shaped relationship. Unemployment tended to be lowest in countries that were either more decentralised or more centralised. Unemployment was highest in countries where the bargaining structures were neither highly centralised or highly decentralised.

The finding was explained in terms that wage increases secured at the industry or sector level are more likely to highlight the trade-off between wage increases and job losses. A number of negative externalities may be associated with the wage bargain secured by a group of well organised workers (OECD 1997:65). Such organised workers may be able to secure relatively large wage increases but at the cost of higher inflation, job losses for some workers, a reduced ability of the unemployed to secure jobs and overall the consequence will be higher rates of inflation and unemployment.

In a centralised wage determination system with strong co-ordination the national parties would set wages that were compatible with national macroeconomic policy objectives. The negative externalities associated with industry and sector wage bargains can be internalised to generate more moderate wage outcomes since the interests of both well organised and poorly organised workers, and high and low productivity sectors can all be taken into account in the wage bargain. Union power could be counterbalanced through wage caps and a strong degree of

equity in the wage structure and wage outcomes could be ensured.

Through either competitive pressures or centralised co-ordination lower wage outcomes are generated in either decentralised wage determination systems or in centralised wage determination systems. Intermediate wage determination systems lack both the countervailing competitive pressure or the countervailing centralised co-ordination. In turn the model assumes that lower wage outcomes are associated with improved macroeconomic performance.

In a review of the Calmfors and Driffil model the OECD (1997) were critical of both its theoretical and empirical content. The level of bargaining is only one dimension of many with other characteristics including the degree of unionisation, the coverage of bargaining and the degree of co-ordination. The linkages between bargaining and labour market outcomes are far more complex and dependent on intervening variables than is suggested by the Calmfors and Driffil analysis. The statistical analysis across countries classified according to different bargaining systems provided no consistent or significant outcomes between bargaining and labour market outcomes.

While the OECD (1994) 'Jobs Study' supported decentralised bargaining as being more in tune with setting wages according to market conditions and allowing for a more market orientated allocation of labour according to wage differentials, the case that decentralised bargaining, such as workplace bargaining, improves macroeconomic performance remains weak. In their analysis of the stagflation experience of the 1970s and 1980s across OECD economies, Rowthorn and Glyn (1991) argued that it was those

countries with strong centralmost countries corporatist arrangements that had a demonstrated superior macroeconomic record.

### **Globalisation, Trade Liberalisation and Workplace Agreements**

The process of political and economic change across countries raises the spectre of whether there is a convergence in bargaining arrangements that is being determined by the pressures of competitiveness. From the rapidly industrialising counties of Asia through to the emerging new market economies of Eastern Europe there is no emerging trend. Even in older liberalised economies like the USA the dominant characteristic of bargaining is its divergence with an extensive unionised sector with plant level collective bargaining existing alongside an extensive non union sector with different forms of representation and bargaining from the informal through to extensive forms of employer participation and representation (Wheeler 1987). Even in the unionised sector there are considerable differences between industry and national unions on the one hand and independent and workplace centred unions on the other hand (Jacoby & Verma 1992).

In South East Asia, the story is one of diversity in bargaining structures. Benson and Debreux (2000) indicate that in Japan, post Asian crisis, while there has been a strong push by employers for a more market driven approach to bargaining, many of the traditional bargaining structures and norms have held. In a study of a new enterprise zone in China, Zuy and Warner (2002) argue that even though there are strong pressures for workplace change that the union movement and the State remain active players in the

bargaining process. Indeed, the strong and centralised role of the state was present in many countries including Singapore, where the government set wage guidelines for bargaining (Chew & Beng 2001) to South Korea where there remained a strong State presence in a tripartite incomes policy arrangement (Lee 2002). In their review of developments in bargaining arrangements across seven Asia-Pacific economies, Bamber et al (2002) found that there was no evidence of convergence in bargaining arrangements despite pressures and policies towards trade and market liberalisation in many of the economies.

In a study of Canadian based corporations operating in North America Hutchins (2002) found that instead of adopting one type of model to bargaining and employee relations, there were within company differences across plants and locations with respect to human resources practices. The companies tended to adapt to local conditions and local arrangements.

In the newly emerging Eastern European economies the impression is one of divergence of bargaining arrangements. In a review of developments Casale (1998) reported on the durability of forms of State participation and control, including centralised collective bargaining, and the presence of national collective incomes policy arrangements in many countries including the Czech Republic and Hungary. Other economies that have undergone a transformation away from nationalisation towards private ownership of the means of production and market exchange still exhibited strong collective and centralist elements (Brainnie 2002).

### **Further Research and Developments**

The practice and relevance of workplace agreement making is being challenged by a

number of developments. First, and as already mentioned, there is the spread of fragmented forms of employment associated with outsourcing and the pursuit of numeric and functional flexibility by employers (Abraham 1998; Allen & Henry 1996). The growth in part-time work, temporary work and contract work means that the link between employer and employee is more tenuous. Such workers are generally difficult for unions to organise and have limited voice within enterprises. There is a concern that moving towards decentralised bargaining risks alienating or excluding workers who face difficulties in being represented in the bargaining process (Hoffman 1997; Supiot 2001). In some cases it may be a whole strata of workers, with, for example, many women workers under the workplace bargaining arrangements in Australia being potentially excluded since they work under part-time or temporary employment arrangements (Sullivan et al 2003).

Second, and associated with this, is the break down in the employment relationship associated with the outsourcing of labour through agencies and through the use of sub-contractors. Such workers are in an ambiguous position regarding their status and are often located outside of the formal bargaining framework (Abraham 1998).

Third, new bargaining units that seek to develop regional and industry norms, are emerging, especially in Europe (Supiot 2001).

Fourth, while there is a trend towards individualism in bargaining there is not as yet, effective forms of representation and voice outside of trade unions. The position of workplace and enterprise based non-union bargaining units will always be compromised and questioned with respect to their legitimacy and independence

(Visser 2001). More decentralised bargaining places a strain on the financial resources of trade unions and strengthens the power of workplace managers (Locke & Kochan 1998).

Fifth, there is interest in the fragmented and more individualised bargaining regimes emerging in the public sectors of many countries as the forces of privatisation, corporatisation and contracting out have had a profound impact not only on employment arrangements and collective representation, but on the bargaining process (Fairbrother et al 2002).

Sixth, there will remain an interest as to how the effects of trade liberalisation and economic transformation impact on bargaining arrangements. In particular whether there is an enduring trend towards convergence in bargaining arrangements (Bamber et al 2002).

The final issue is the tension between the trend towards more decentralised bargaining and individualism in industrial relations as opposed to the development of national and international norms in employment. No where is this more apparent than in the EU with its social framework (Supiot 2001; Visser 2001).

## Selected References

- Abraham, K. (1998) "Restructuring the Employment Relationship: the Growth of Market Mediated Work Arrangements", in F. Ackerman, N. Goodwin, L. Dougherty and K. Gallagher (Editors), *The Changing Nature of Work*. Washington DC: Island Press, pp. 185-189
- Allen, J. and N. Henry. (1996) "Fragments of Industry and Employment: Contract Service Work and the Shift Towards Precarious Employment", in R. Crompton, D. Gallie and K. Purcell (Editors), *Changing Forms of Employment: Organisations, Skills and Gender*. London: Routledge, pp 65-82.
- Armstrong, P., J. Goodman and J. Hyman. (1981) *Ideology and Shop Floor Industrial Relations*. London: Croom Helm.
- Bamber, G. and R. Lansbury. (1987) "Studying Industrial Relations", in G. Bamber and R. Lansbury (Editors), *International and Comparative Industrial Relations*. Sydney: Allen and Unwin, pp. 3-30.
- Bamber, G., C. Leggett and P. Ross. (2002) "Changing Employment Relations in Seven Asia Pacific Countries", in T. Hamani (Editor), *Universal Wisdom Through Globalisation*. Tokyo: Japan Labour Institute, pp. 287-298.
- Benson, J. and P. Debroux. (2000) "Japanese Trade Unions at the Cross Roads—Dilemmas and Opportunities Created by Globalisation", in C. Rowley and J. Benson (Editors), *Globalisation and Labour in the Asia Pacific Region*. London: Frank Cass, pp. 114-132.
- Booth, A. (1995) *The Economics of Trade Unions*. Cambridge: Cambridge University Press.
- Brainine, M. (2002) "Algerian Employment Policies and Practice: An Overview", *International Journal of Employment Studies*, Volume 10, Number 1, pp. 133-152.
- Brown, W. and P. Ryan. (2003) "The Irrelevance of Trade Union Recognition? A Comparison of Two Matched Companies", *Australian Journal of Labour Economics*, Volume 6, Number 3, pp. 383-408.
- Calmfors, L. and J. Driffl. (1988) "Bargaining Structures, Corporatism and Macroeconomic Performance", *Economic Policy*, April, pp. 15-61.

- Casale, G. (1997) "Recent Trends and Issues in Industrial Relations in Central and Eastern Europe", in J. Brady (Editor), *Industrial Relations in the Market Economy*. Dublin: Oak Tree Press, pp. 1-14.
- Chew, R. and C.S. Beng. (2001) "Evolution of the Wage System in Singapore", *International Employment Relations Review*, Volume 7, Number 2, pp. 27-43.
- Cully, M., S. Woodland, S. O'Reilly and G. Dix. (1999) *Britain at Work: the 1998 Employee Relations Survey*. London: Routledge.
- Deery, S. and D. Plowman. (1991) *Australian Industrial Relations*. Third Edition. Sydney: McGraw-Hill.
- Fairbrother, P., M. Paddon and J. Teicher. (2002) "Corporatisation and Privatisation in Australia", in P. Fairbrother, M. Paddon and J. Teicher (Editors), *Privatisation, Globalisation and Labour*. Sydney: Federation Press, pp. 1-24.
- Freeman, R. (1989) *Labour Markets in Action*. Cambridge, Mass: Harvard University Press.
- Freeman, R. and J. Medoff. (1984) *What Unions Do?* New York: Basic Books.
- Hoffman, R. (1997) "Collectivism and Individualism", in D. D'Art and T. Turner (Editors), *Collectivism and Individualism: Trends and Prospects*. Dublin: Oaktree Press, pp. 1-10.
- Hutchins, K. (2000) "The Effects of Global and Regional Political and Economic Trends on Employment Relations in Canadian Organisations", *International Employment Relations Review*, Volume 6, Number 1, pp. 35-60.
- Jacoby, S. and A. Verma. (1992) "Enterprise Unions in the United States", in M. Bagnanna and M. Kleiner (Editors), *Labour Market Institutions and the Future of Trade Unions*. Cambridge: Blackwell, pp. 137-158.
- Lansbury, R. and D. Macdonald. (1992) "Issues in Workplace Industrial Relations", in R. Lansbury and D. Macdonald (Editors), *Workplace Industrial Relations: Australian Case Studies*. Melbourne: Oxford University Press, pp. 2-28.
- Lee, B.-H. (2002) "Labour Unions in the Evolution of Social Dialogue: the Korean Experience", *International Employment Relations Review*, Volume 8, Number 1, pp. 47-55.
- Locke, R. and T. Kochan. (1998) "Conclusions: the Transformation of Industrial Relations? A Cross National Review of the Evidence", in F. Ackerman, N. Goodwin, L. Dougherty and K. Gallagher (Editors), *The Changing Nature of Work*. Washington DC: Island Press, pp. 231-234.
- Macdonald, D. and J. Burgess. (1999) "Outsourcing, Employment and Industrial Relations in the Public Sector", *Economic and Labour Relations Review*, Volume 10, Number 1, pp. 36-55.
- Macpherson, D. and J. Stewart. (1990) "The Effects of International Competition on Union and Non Union Wages", *Industrial and Labor Relations Review* Volume 43, April, pp. 435-446.
- Millwald, N. and M. Stevens. (1986) *British Workplace Industrial Relations 1980-1984*. Aldershot, UK: Gower.
- Morehead, A., M. Steele, M. Alexander, K. Stephen and L. Duffin. (1997) *Changes at Work: the 1995 Australian Workplace Industrial Relations Survey*. Melbourne: Longman.
- Nash, J. (1950) "The Bargaining Problem", *Econometrica*, Volume 18, pp. 155-162.



von Neumann, J. and O. Morgenstern.  
(1944) *The Theory of Games and Economic Behaviour*. Princeton: Princeton University Press.

OECD. (1994) *The Jobs Study*. Paris: OECD.

OECD. (1997) *Employment Outlook*. Paris: OECD.

Rowthorn, R. and A. Glyn. (1991) "The Diversity of Unemployment Experience Since 1973", in S. Marglin and J. Schor (Editors), *The Golden Age of Capitalism*. Oxford: Clarendon Press, pp. 218-266.

Sullivan, A., G. Strachan and J. Burgess. (2003) "Women Workers and Enterprise Bargaining", in J. Burgess and D. Macdonald (Editors), *Developments in Enterprise Bargaining in Australia*. Melbourne: Tertiary Press, pp. 158-175.

Supiot, A. (2001) *Beyond Employment: Changes in Work and the Future of Labour Law in Europe*. Oxford: Oxford University Press.

Visser, J. (2001) "Industrial Relations and Social Dialogue", in P. Auer (Editor), *Changing Labour Markets in Europe*. Geneva: ILO, pp. 178-242.

Wheeler, H. (1987) "Management Labour Relations in the USA", in G. Bamber and R. Lansbury (Editors), *International and Comparative Industrial Relations*. Sydney: Allen & Unwin, pp. 57-79.

Ying, Z. and M. Warner. (2002) "Human Resource Management in China's Frontier Special Economic Zone: A Study of Selected Enterprises on Hainan Island", *International Journal of Employment Studies*, Volume 10, Number 1, pp. 75-104.

*John Burgess*  
*Newcastle Graduate School of Business*  
*University of Newcastle*  
*Australia*  
*John.Burgess@newcastle.edu.au*